RELATIVE EFFICIENCY OF MACHINE LEARNING MODULES FOR PREDICTION OF MILK YIELD AS A TOOL FOR GENETIC EVALUATION IN BOVINES

Thesis

Submitted to Guru Angad Dev Veterinary and Animal Sciences University
in partial fulfillment of the requirements for the degree of

MASTER OF VETERINARY SCIENCE
in
ANIMAL GENETICS AND BREEDING
(Minor Subject: Animal Biotechnology)

By

Revathy T
(L-2021-V-03-M)

Department of Animal Genetics and Breeding
College of Veterinary Science
©Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana – 141004

2024

CHAPTER I

INTRODUCTION

India is the largest milk-producing country in the world with a production of 230.58 million tonnes. The per capita availability of milk is around 459 grams per day in 2022-23 (BAHS, 2023). The total population of cattle and buffalo in India is 193.46 and 109.85 million, respectively (Livestock Census, 2019). Out of the total milk produced in the country, cattle contribute 51.91% (Crossbreed 29.81%, Indigenous 10.73%, Nondescript 9.51% and Exotic 1.86%), and buffalo shares about 44.81% including both Indigenous (31.94%) and Non-descript (12.87%). Additionally, there is a 3.69% increase in buffalo milk production compared to the previous year (2021-22). Milk production from exotic/crossbred cattle has risen up by 3.75%, and that from indigenous/non-descript cattle has seen a 2.63% increase compared to the previous year. Among 15 major milk-producing states, Punjab contributed a yield rate of 13.49kg/day/animal in exotic/crossbreed and 8.68 kg/day/animal in indigenous/non-descript animals in the year 2022-23. In the per capita availability of milk, Punjab holds the highest value of 1283 grams/day/head among all the states, including the All-India average (BAHS, 2023).

Dairy production is one of the main sustenance factors for the rural economy of India, alleviating rural poverty, in the form of providing better nutrition and quality protein, warranting food security for millions of rural households; enhancing economic growth requires optimizing selection and culling decisions, particularly in dairy animals. Predicting milk production is a valuable tool for dairy farmers as it enables them to make informed decisions regarding herd management, ultimately improving overall efficiency and productivity on the farm (Dematawewa et al., 2007; Grzesiak et al., 2006). The profitability of a dairy herd depends on the overall lifetime production of the cows in the herd. In India, the breeding program is mainly appraised based on the first 305-days milk yield for how it is necessary to record the data of daily milk yield. The daily milk yield (DMY) holds significant importance in the identification of cows that have not achieved their anticipated milk production trajectory due to factors such as illness or other welfare concerns. On the contrary, daily milk yield recording is tedious, laborious, expensive/extravagant, and practically unachievable under field conditions.

Lactation length can be defined as the period of milk produced from the date of calving to the date of drying or cessation of milk production excluding the first 4 to 5 days of colostrum. The standard lactation length for cattle and buffalo is said to be 305 days. This

parameter plays a major role in the selection and early prediction of a low-producing animal for timely culling during the production process and for genetic evaluation. Prediction of 305-days milk yield is very important for better herd management and to overcome the economic loss. Over a period, several conventional methods were suggested for predicting the 305-days milk yield. In which, the regression model has been the only way to build the prediction of cow/animal-related factors. The regression models are the core algorithms of both conventional statistics and Machine Learning. The former category of algorithms depends on standard linear regression techniques, such as stepwise multiple linear regression, segmented multi-phase regression, and generalized linear regression (Rubinfeld 1998). However, the practical application and prediction power of the above regression methods is hardly bounded due to their structural restriction and the quality of available data. In consequence, machine learning algorithms are progressively being preferred for flexible assessment and higher precision.

The main criteria used in dairy cattle breeding programs are milk yield and milk composition. The majority of cows are milked twice a day either by hand or by machine. The standard milk yield recording is done every month during lactation to get the official milk records for the herd book. The production of milk and its composition provides the farmers with their revenue. For their economy to function, milk yield must be measured or predicted with accuracy (Fernandez et al., 2007).

Milk yield selection has predominantly focused on the 305-days milk yield, as noted by Ptak and Schaeffer (1993), Seyedsharifi et al. (2008), and Bilal and Khan (2009). An alternative genetic evaluation method involves directly utilizing test-day (TD) milk yields, as proposed by Kaya et al. (2003) and Tailor and Singh (2011). The TD models analyze individual records of cows on specific test days, diverging from the traditional reliance on 305-days milk yields (Pool and Meuwissen, 1999). The primary goals of implementing selection based on test-day (TD) records were to minimize recording expenses and enhance the precision of genetic evaluation (Nigm et al., 2003). The generation interval can be shortened and selection intensity can be increased with early assessments of the sire's breeding value by prolonging lactations in progress leading to more genetic advancement. The farmer can use this early knowledge to determine whether or not to keep cows for breeding. Additionally, it facilitates the distribution of resources, such as feed, for both a herd and a single cow (Khan et al., 2005)

Machine Learning is a subset of Artificial Intelligence that enables the computer to automatically create statistical models; through which it can generate predictions, based on

what they have learned. These algorithms are used to reduce the dimensionality of the calibration data without compromising the predictive power of the module. Predictive models about the behaviours, health, and productivity of cows hold significant value within the contemporary dairy industry. These models possess the capability to aid in decision-making processes on farms and the marketing of products.

The main objective of animal breeding is to enhance the performance of each generation of animals. To achieve this it is crucial to identify and select individuals based on their breeding values, which represent their genetic merit. The process of determining these breeding values is a data-intensive process that plays a pivotal role in enhancing productivity. Traditionally statisticians and breeders have relied on established methods to estimate these values leading to advancements in the quality and productivity of livestock. However recent studies (Hamadani et al., 2022) have explored the use of machine learning algorithms for prediction of genetic merit in dairy sheep, which have exhibited promising results.

The genetic evaluation of dairy animals heavily depends on the milk yield records of the animals and their relatives viz. dams and daughters. The recording of the milk yield in the field daily has been a part of routine operation at institutional farms but has never been attractive to the farmers as they do not perceive an immediate benefit from that. Thus, the scope of the animals from which a genetic evaluation can be done in the country has been greatly limited. The evaluation of sires using a test-day milk yield model offers adequate accuracy, primarily attributed to a higher number of measurements per daughter compared to the 305-days model based on a single lactation record. This heightened accuracy may stem from avoiding bias introduced by discarding incomplete records, which are not utilized in predicting the 305-days milk yield. Furthermore, the test-day model allows for adjusting more comprehensively for detailed fixed effects, providing a more nuanced assessment. This model accommodates environmental variations specific to individual test-day milk yields and captures genetic effects associated with individual animals. In the context of sire evaluation, the test-day milk yield model minimizes residual variance by leveraging all available test days of a sire's daughters, thereby offering more information per sire. It also identifies and adjusts for a wider range of environmental effects in the genetic evaluation model.

Gandhi and Gurnani (1991) observed that selecting bulls based on production efficiency traits would result in more significant genetic improvement within the herd. This is attributed to the traits' high heritability and favourable correlation with lactation traits. The foundation of any breed improvement initiative relies on the careful selection of genetically superior sires. The effectiveness of such programs hinges on the timely and

precise evaluation of young bulls. Various approaches, including 'least squares' (LSQ), 'simple regressed least squares' (SRLS), 'best linear unbiased prediction' (BLUP), and the 'restricted maximum likelihood method' (REML), have been employed by different researchers for sire evaluation.

The field progeny testing program leverages the test day milk yield recording system for the estimation of 305-days milk yield of the animals based on regression/ lactation curve models. However, the test day milk yields are rigid for the days on which the milk yields are to be recorded and prone to errors in case of any deviation. Machine learning can provide a plausible alternative to the traditional methods of milk yield estimation while allowing flexibility on the days of milk yield recording. The achievement of breeding goals within a breeding program revolves around the precise identification of genetically superior animals to become the progenitors of the next generation. This practice is fundamental to the enhancement of genetic traits and the overall progress of the breeding program. However, one critical aspect of this process involves accurately predicting the breeding values of potential selection candidates, and this step often proves to be laborious and computationally demanding. Therefore, the present study intends to develop a biological-based, machine learning model to predict the 305 day milk yield with the following objectives:

1. To train the machine learning model for prediction of 305-days milk yield of the cattle and buffaloes.
2. To evaluate and compare the efficiency of trained model for the prediction of 305-days milk yield against conventional methods.
3. To compare the breeding value of the animals from the observed and predicted 305-days milk yield records.

CHAPTER II

REVIEW OF LITERATURE

Dairy farming, which primarily focuses on producing milk involves a range of practices aimed at maximizing the quantity and quality of milk from dairy animals. When it comes to genetic selection processes both the composition and yield of milk are factors to consider. These traits provide insights into how productive and high quality dairy animals (Erfani-Asl et al., 2015). It emphasizes the importance of maintaining, accounting and recording systems to ensure effective management and financial sustainability. By keeping records farmers can analyse patterns, identify areas for improvement and make informed decisions to boost overall productivity. The integration of accounting and recording systems in dairy farming plays a role in breeding management by enabling the analysis of individual and herd performance, optimizing genetic selection methods, and ensuring economic viability, in the pursuit of enhanced milk production.

2.1 Various milk yield recording systems

Milk yield recording systems serve as indispensable equipment in the dairy industry, providing a comprehensive assessment of an individual cows' milk production performance. By meticulously monitoring and analysing milk yields, these systems empower dairy farmers with vital insights that increase past mere quantitative outputs. The records generated turns into a backbone for informed decision-making throughout crucial domains, which includes breeding techniques and herd management. Milk recording structures make a contribution to the identity of high-performing cows, allowing farmers to make strategic decisions on mating and genetic selection. This, in turn, enhances the overall genetic potential of the herd, fostering a lineage of cows with superior milk-producing attributes. Furthermore, milk recording systems enable farmers to implement specific herd management practices. By identifying outliers and monitoring traits in individual cow performance, farmers can proactively address health problems, optimize reproductive cycles, and implement preventive measures. This holistic approach complements the overall productiveness and durability of the dairy herd by accounting the milk in various patterns in form of Lactation milk yield (LMY), 305 day milk yield (305DMY), test day milk yield (TDMY).

2.1.1 Daily milk yield

The milk yield on each day is the sum of the milk yields of all milking's within a day (morning and evening), which are recorded on that specific date and expressed in kilograms per day. For dairy farmers, daily recording of milk yield data ultimately poses logistical and practical challenges. Nevertheless, overcoming these obstacles is crucial, since fast and reliable data is the foundation of dairy operations that succeed, allowing farmers to maximize herd productivity, improve genetic selection, and increase overall herd size. It also emphasizes the necessity of employment, automation and cutting-edge technologies in contemporary dairy farming to expedite data collecting and lessen some of these challenges.

2.1.2 Lactation milk yield

In the world of dairy farming, lactation yield (LY) is the most important trait which is accounted for, the amount of milk produced by a dairy cow over the period from calving to drying off. It serves as a pivotal parameter that demands daily recording, as its accuracy and consistency are essential for effective selection and management in the dairy industry. Daily monitoring of lactation yield is essential for optimizing operations and ensuring selection leading to more productive herds. However, recording of daily milk yield data poses a number of serious difficulties. The amount of milk produced fluctuates not only from day-to-day measures but also depends on a wide range of effects, such as the health of the cow, the environment it lives in, and the nourishment it consumes. To ensure measurement of milk output it is necessary to accurately record these complexities

2.1.3 305-days milk yield

The 305 day milk yield (305DMY) represents the amount of milk that a cow produces during the lactation period from the first day, until the 305th day. It serves as a crucial metric for selecting and culling cows in the production process and plays a significant role in individual genetic evaluation (Kong et al., 2018). In dairy breeding, the 305DMY offers an extensive evaluation of a cow's productivity over a standard period and has historically been used to select for milk yield estimation (Ptak and Schaeffer, 1993; Seyedsharifi et al., 2008; Bilal & Khan, 2009; Bilal et al., 2009). However, it is very costly as well as difficult to record a 305-days supply of milk yield in Indian scenario. As a result, the alternative strategy evaluating sires based on test day

milk yield (TDMY) has received more attention. The accuracy of 305-days yields depends on the number of test-days (TDs), the number of days between tests, and the methods of estimating 305-days yields (Kaya et al., 2003).

A transition from the conventional 305-days milk yield (305DMY) recording to the test-day milk yield recording brings numerous benefits. By replacing 305DMY, test-day milk yield effectively removes environmental variations in phenotypic milk yield data (Cobuci et al., 2005). Unlike the 305-days approach, test-day milk yield accounts for specific environmental effects for each production record (Visscher and Goddard, 1995), thereby allowing for a more accurate evaluation of cows through the utilization of a greater number of records per cow (Rekaya et al., 2001). Additionally, the test-day milk yield method enables a more precise estimation of genetic and permanent environmental effects on milk yield by considering specific factors influencing each production record (Seyedsharifi et al., 2008).

Nevertheless, the use of directly measured TD milk production is an alternative and becoming a more and more popular method for genetic evaluation (Kaya et al., 2003; Tailor & Singh, 2011). The shift in methodology has been more widely recognized recently and has many benefits. Test-day data has long been recommended to be used to evaluate dairy cattle rather than relying solely on complete or 305-days lactation records (Guo & Swalve, 1995).

2.1.4 Test day milk yield

Test day milk yield (TDMY) is a measurement of the milk produced by a cow over 24 hours (Santos et al., 2013; Schaeffer & Jamrozik 1996). Test-day milk yields track the amount of milk a cow produces on particular days during its lactation cycle rather than depending solely on the cumulative 305-days amount. This method offers a more comprehensive and nuanced picture of a cow's milk production, making it possible to assess the genetic potential with greater accuracy. Swalve (2000) reported, the test-day models are more adaptable in managing records from various recording schemes; they have gained a lot of interest in the past few decades.

The use of a test day model (TDM) for the genetic evaluation of milk production attributes in dairy cattle has received a lot of attention recently. A TDM's benefits include more accurate correction for environmental factors that are pertinent to every test day (TD) record. The main obstacle that greatly impacts the sire evaluation

programme is the lack of adequate data, which institutional herds cannot supply on their own. Thus, gathering and using field data becomes essential for genetic advancement.

Santos et al. (2013) concluded that there would be indirect benefits in 305MY and Lactation Length (LL) from the selection criteria that make use of test day milk yield (TDMY). On the other hand, direct selection for this trait would yield the greatest response to selection for 305 MY.

Swalve (2000) reported that instead of combining records, test-day (TD) models use data from specific test days to estimate lactation production. Through frequent genetic examinations using the most recent data, they can minimize the generation interval compared to solely using records of full lactations. Test day records offer flexibility in managing data from various recording schemes and can support a variety of recording schemes, from significant data collection that involves only a few data points per lactation to daily recording of milk yield.

Bilal and Khan (2009) highlighted that the traditional 305-days lactation model might inaccurately overestimate incomplete lactations as completed ones as well as overlook adjustments for environmental variations throughout the 305-days lactation period. Recently, the use of test-day models has allowed for more accurate genetic examination to be done with less expense and better at managing records from various recording schemes. Moreover, the prevalent approach of solely aggregating TD (test-day) records into 305-days lactation records undermines their endeavors to enhance evaluation models. This counteracts the efforts aimed at minimizing environmental influences by employing meticulously accurate correction factors and refining the definition of contemporary groups.

The part lactation yields were utilized to project 305 days yield, before popularization of TDMs. Projection methods commonly adopt a standardized lactation curve shape for cows within the same calving season, age category, and herd level. However, this approach tends to underestimate the 305-days yields for highly persistent cows from early test-days, while overestimating yields for less persistent cows (Mrode et al., 2002).

Briefly defined, TD models: (i) increase the precision of genetic evaluation; (ii) provide modelling more effectively than other models; (iii) directly take into consideration all genetic and environmental effects on a TD basis; (iv) represent that

extending part lactation is no longer required. and (v) suggest that the expenditure of milk recording may also be decreased by having longer intervals between milk recording and collecting fewer samples of milk (Swalve, 1995; Kaya et al., 2003; Bilal & Khan, 2009; Kaygisiz 2013). Nevertheless, a TD model is unable to compensate for the accuracy loss resulting from utilizing less TD or inaccurate recording (Wiggans and Goddard, 1997). The primary goals of the TD selection process were to reduce the cost of recording and improve the precision of genetic assessment (Nigm et al., 2003).

However, it is not practical to attain data at close intervals by taking into account various other factors. To conquer this difficulty, several factors have been reported by many researchers in different ways viz. Test Day Milk Yield in the form of Weekly (Sahoo et al., 2016), Fortnightly (Sahoo et al., 2014), Monthly and Bimonthly (Sahoo et al., 2015).

Bilal et at. (2009) concluded that by recording seven to ten test days (TD) every month, it is possible to calculate 305 days of lactation. Breed, herd management, management group within a herd, days in milk, age at calving and test, pregnant status, and number of milking times per day are some of the factors that determine TD yields. Two distant TD yields would give more information than those that are close and highly connected. The use of TD data would allow the use of information from lactations with extensive periods between milk recordings because the calculation of yields for unrecorded intervals would not be required.

Sahoo et al. in 2014, using both fortnightly and weekly test day milk yield records during the first lactation of 961 Murrah buffaloes, consistently found that the Polynomial Regression Function (PRF) model was the best mathematical model for predicting test day milk yields.

Chakraborty et al. (2020) found that the accuracy of bimonthly test day yield (BMTY) recording is comparable to that of lactation yield and monthly test day records. The accuracy of Selection (rIH) for odd test day records (TD1 + TD2 + TD3 + TD4 + TD5 + TD7 + TD9) was 0.686, and for even test day records (TD2 + TD4 + TD6 + TD8 + TD10), it was 0.652. In contrast, the direct selection accuracy was 0.566. Notably, the alternate method of test day records of the first lactation demonstrated higher selection accuracy than direct selection (0.566). Researchers suggested that

opting for bimonthly recording of test day milk yield data could be a more cost-effective alternative for bull selection in Murrah buffaloes.

Kokate et al. (2014) reported in their study of 1345 Karan Fries cows that the regression coefficients demonstrated a steady rise up to the BMTDMY3, followed by a consistent decline thereafter until BMTDMY-7, indicating a continuous trend in the test day milk yield. The predictive accuracy (R^2) for 305-days milk yield, relying on individual bimonthly test day milk yields, ranged from 19% (BMTDMY-1) to 59% (BMTDMY-4). Notably, higher prediction accuracy was observed in the mid-phase of lactation for bimonthly test day milk yields.

Out of the above, test-day milk yields systems are being used more and more to estimate milk production in dairy cattle. One of the selection criteria for dairy cows with the potential to make a lifetime profit is the capacity to predict FL305DMY with the greatest degree of accuracy utilizing test day milk yields during the early stages of lactation (Gandhi & Gurnani, 1988; Kannan & Gandhi 2006; Dongre et al., 2012; Kumar et al., 2019).

2.2 Effects of non-genetic factors on test day milk yield

The primary goal of every breed improvement initiative is to achieve genetic advancements through the selection of superior sires and dams. The test-day milk yield model considers environmental variations specific to individual test-day milk yields and genetic effects linked to individual animals. Gupta et al. (2020) reported that assessing the impact of non-genetic factors on both test-day and full lactation milk yields is valuable for predicting selection responses. Moreover, this information can be utilized to enhance breeding programs for overall improvement and increased genetic gains in organized farms.

2.2.1 Season of calving

Season has been identified as a significant environmental factor influencing cattle performance. Considering the considerable variation in climatic conditions throughout the year, differences are more pronounced between seasons and less within a particular season.

Khosla et al. (1984) and El-Arian (1986) found that the season of calving (SOC) significantly influenced (P≤0.05) all monthly test-day milk yields in Murrah buffaloes.

Singh et al. (2020) based on the results of the statistical analysis revealed that the season of calving in 191 Murrah buffaloes had a highly significant (P<0.01) impact on MTDMY1, MTDMY2, MTDMY3, and MTDMY4, and a significant (P<0.05) effect on MTDMY5, MTDMY6, and MTDMY7. The researchers also reported buffaloes that calved in the winter season consistently exhibited the highest individual monthly test day milk yields across most test days, with the exception of MTDMY2 and MTDMY3. In contrast, the lowest monthly test day yields were observed in buffaloes calved during the autumn season, except for MTDMY1, MTDMY9, and MTDMY10.

According to Katneni (2007), the season of calving significantly affected MTDMY5 to MTDMY9, while Patil (2011) found that the season of calving significantly affected MTDMY2, MTDMY5, and MTDMY7 in Murrah buffaloes, whereas there was no significant effect observed on the remaining MTDMYs. Sharma et al. (2017) estimated that the season of calving had significant effect (p<0.05) on on 3rd and 9th test day milk yield among the 11 test day milk yield.

Yoon et al. (2004) demonstrated that cows giving birth in spring and winter exhibited higher milk production compared to those calving in summer. Similarly, El-Tahawy and El-Far (2010) found that milk yield reached its peak during autumn and winter, surpassing the yields observed in spring and summer.

However, Singh and Yadav (1987), Sahana (1993), and Rana (2008) found that the season of calving had no significant impact on traits related to monthly test-day milk yields in Nili-Ravi buffaloes, Sahiwal (Kumar et al., 2021) and Jersey cows (Ertuğrul et al., 2018).

2.2.2 Period of calving

The entire MTDMYs in Murrah buffaloes were significantly impacted by the period of calving (POC), according to Khosla et al. (1984); El-Arian (1986); Patil (2011); Jamuna (2012), Kumar (2012); Singh (2014); Rohit et al. (2020); Gurmessa and Melaku (2012); Jónás et al. (2016) and Stanton et al. (1992).

In a study conducted by Singh et al. (2020) it was found that the monthly test day milk yields for MTDMY2, MTDMY3 and MTDMY7 had a effect on the period of calving (p<0.01) while MTDMY1, MTDMY4, MTDMY5 and MTDMY8 showed a significant effect (p<0.05).

Sharma et al. (2017) also estimated that the period of calving had a highly significant on the second test day and a significant on the third test day in Jaffarabadi buffaloes (p<0.01 and p<0.05 respectively).

Additionally, Dass (1995) observed that except for the test day milk yield during the first lactation of Murrah buffaloes at day 155th day, after calving all other test day milk yields were substantially affected by the period of calving.

In a study conducted by Geetha (2005) it was discovered that there was an impact of POC, on the test day milk yields. This impact was found to be significant on days except for the 215th and 305th days.

Tailor and Singh (2011) also observed an significant effect (p≤0.05) on the 1st, 4th, 6th and 7th MTDMYs. They further noted highly significant impact (p≤0.01) specifically on the 9th test day milk yield.

Kumar et al. (2012) reported the findings indicating a significant effect of POC on the 9th test-day milk yield.

On another note, Singh (2014) observed a POC effect (p <0.01) on all monthly test day milk yields except for the ninth one.

In contrast to these findings Sudhakar et al. (2013) found no impact of period of calving on the test day milk yield in Holstein cows.

2.2.3 Parity

The influence of parity, on test day milk yield was found to be significant (Ertuğrul et al., 2018). Erdem et al. (2010) also concluded that parity has a significant impact on test day milk yield (TDMY). The study revealed that the first parity had the lowest TDMY, with a statistically significant difference (P= 0.011). Additionally, the study found that the third parity exhibited the highest lactation milk yield (LMY), showing a linear decrease in subsequent lactations (P= 0.009).

On the contrary Kumar et al. (2021) found no significant influence of parity, on test day milk yield. However they observed that during specific parities, the third parity had the highest test day milk yield while the fourth parity recorded the lowest.

2.2.4 Dry period

The dry period refers to the period from the date of drying (stop of milk production) to next calving. It is crucial to minimize this non-productive period to maximize profits.

Various studies have examined the impact of the period of calving on dry period in different buffalo breeds. Thiruvenkadan (2011), Singh et al. (2011), Barman et al. (2012) and Jakhar et al. (2016) have all highlighted its influence. However, Kumar (2000) found no significant effect in Murrah buffaloes. Conversely Jakhar et al. (2016) emphasized a influence of the calving season on the dry period for Murrah buffaloes at a highly significance level (p<0.01). On the hand, Suresh et al. (2004) and Barman et al. (2012) reported that there was no effect of season of calving and dry period, in Murrah buffaloes.

2.2.5 Service period

Service period defined as the period between calving and the subsequent successful conception, typically allows for an optimal duration of 60 days. This duration is influenced by factors, including management practices, environmental conditions, the regularity of oestrus (heat) and the number of services required for conception (Jakhar et al., 2017).

Several studies have examined the relationship between the calving period and service period in Murrah buffaloes. Research conducted by Suresh et al. (2004) Godara et al. (2004) and Wakchure et al. (2008) revealed significant impact of the period on the service period. However Kumar et al. (2005) found no effect of the period on service period across different buffalo breeds.

Similarly, Dhirendra et al. (2003), Suresh et al. (2004), Kumar et al. (2005) and Jakhar et al. (2016) discovered that the season of calving had significant influence on the service period, in Murrah buffaloes.

2.3 Genetic parameter estimation

Estimation of genetic parameter in cattle and buffalo based on test day milk yield record is crucial for effective and sustainable dairy animal breeding programs. Accurate genetic parameter estimates play a crucial role in genetic enhancement for future milk yield through the selection of dairy cattle, as highlighted by Aspilcueta-Borquis et al. (2010). By accounting test day milk yield data, researchers can estimate heritability, genetic correlations, and other important genetic parameters. These parameters provide valuable insights into the degree to which milk yield is influenced by genetic factors, enabling breeders to make informed decisions in selecting superior animals for breeding. Accurate estimation of genetic parameter helps in optimizing breeding strategies, leading to the development of animals with improved milk

production efficiency, disease resistance, and overall productivity (Lodhi et al., 2015). This information not only enhances the economic viability of dairy farming but also contributes to the long-term genetic improvement of cattle populations, promoting a sustainable and resilient dairy industry Comparisons were made between genetic evaluations of dairy sires and cows for milk production using 305-days lactation yields and evaluations based on the respective test day yields from those lactations, as studied by Ptak and Schaeffer (1993).

The influence of genetic and environmental factors on 305-days milk yield has been accounted for as a general effect, as noted by Kettunen et al. (1998). However, multiple factors contribute to changes in milk yield over the lactation period. Consequently, instead of relying on the calculation of 305-days milk yield, researchers have considered the use of test day milk yield (Reents and Doop, 1996; Olori et al., 1999; Kaya et al., 2003; Jamrozik et al., 1997; Swalve, 1995). This shift has led to more accurate genetic evaluations, resulting in an increased pace of genetic improvement, as observed by Akbas (1994).

Table 2.1:   Correlations between estimated breeding values for TDMY and 305DMY

Breeds	Estimated breeding value	Authors
British Holstein-Friesian	0.57 to 0.99	Pander et al. (1992)
Dairy cows	0.87 to 0.97	Ptak and Schaeffer (1993)
Friesian cows	0.88 to 0.96	Swalve (1995)
Canadian Holstein cows	0.87-0.97	Jamrozik et al. (1997)
Hariana cows	0.23 to 0.95	Dalal et al. (1999)
Holstein cows	0.89	Rekaya et al. (2001)
Gir cows	0.85 to 0.94	Ledic et al. (2002)
Holstein cows	≥ 0.97	Kaya et al. (2003)
Iranian Holstein cows	0.93 to 0.97	Seyedsharifi et al. (2008)
Brown Swiss Cows	0.66 to 0.86	Çilek and Kaygisiz (2008)
Sahiwal Cattle	0.923 to 0.927	Khan et al. (2008)
Simmental Cattle	0.876	Çilek et al. (2008)

The test-day model incorporates milk yield records collected at different points throughout the lactation period. Instead of extending lactations to 305 days, individual test day milk yields can be employed for the genetic evaluation of dairy sires and cows. Relative amount of genetic variation during lactation is a major factor in the use of TD yield (Seyedsharifi et al., 2008). In contemporary times, the test-day model is utilized for the genetic evaluation of dairy cattle, replacing the 305-days model (Strabel et al., 2005). An alternative method for genetic evaluation involves assessing the genetic potential using test day milk yields directly, as implemented in the models referred to as milk yield models (Firat, 2001).

In Seyedsharifi et al. (2008), the Estimated Breeding Values (EBVs) for test-day (TD) milk yields exhibit a close correlation with the EBVs for 305-days milk yields. This finding aligns also with the outcomes reported by Hammami et al. (2008) and Jamrozik and Schaeffer (1997). The correlations between EBVs for 305-days and TD milk yields, both for cows and sires, were exceptionally high, exceeding 0.92. Specifically within sires, these correlations increased with the number of daughters per sire, indicating improved accuracy. Consequently, this suggests that test-day milk yields could serve as a viable alternative to 305-days milk yield in the genetic evaluation of dairy cattle.

Santos et al. (2013) on evaluating the genetic parameters for 305-days cumulative milk yield (MY305), test-day milk yield (TDMY), and lactation length (LL) using the data from the first lactations of 2816 Guzerat cows selected for milk production in 28 Brazilian herds concluded that the predicted heritability's for MY305 and LL were 0.12 and 0.24, respectively, while the range for TDMY was 0.16 to 0.24. High and favourable genetic correlations were found, with values ranging from 0.51 to 0.99 within TDMY records, 0.81 to 0.98 between each TDMY and MY305, and 0.71 to 0.94 between each TDMY and LL. Based on genetic factors derived from this study researchers reported that test-day milk yield records derived are found to be a useful tool for predicting breeding values for milk production for a Guzerat breed population.

Jensen (2001) reported that the traditional method of displaying aggregated 305-days milk yields for genetic evaluation of dairy cattle has recently been replaced with individual test-day records (TDR). Test-day models (TDM) have several benefits, such as the capacity to take account of the environmental impacts of each test day, the ability

to predict the lactation trajectory for specific genotypes or groups of animals, and the potential for genetic assessments for persistency of production. In addition, test-day models eliminate the need to prolong the lactations of animals that have been culled or that have ongoing records.

Sahoo et al. (2014) concluded that the estimated genetic and phenotypic correlations between the various weekly test day milk yield (WTDMY) and the 305-days milk output varied from 0.19 to 0.98 and from 0.52 to 0.84, respectively. Also demonstrated that all of the WTDY and FL305DMY were significantly affected (p<0.01) by the calving period. These TD yields could serve as the selection criteria for the early assessment and selection of animals, as suggested based on estimations of phenotypic and genetic correlations that were generally higher in the middle segment of lactation.

Dairy animal genetic evaluations are based on studies of the milk produced during the 305-days lactation period in several countries. Therefore, it would be beneficial to put statistical models into practice for forecasting 305-days productivity in subsequent lactations based on past lactations or total lactation yield based on early records. Conventional statistical models have been utilized extensively as prediction tools for a variety of real-life issues. These models include logistic regression, partial least-squares regression, multiple linear regression, stepwise multiple linear regression, linear regression, and projection pursuit regression (Kumar et al., 2019).

The analysis of TD records entails the analysis of repeated records on an individual. Swalve (2000) divided TDM in the following classes: Two-step TDM, One-step TD Models with Fixed Regression, Random Regression-TDM, Multiple-trait (MT) TDM, Character process models (CPM), Covariance function models and splines models, etc. Jensen (2000) reported that it should be evident that a wide range of models can be suggested for TDR, and at the moment, there is no consensus on which model best matches the lactation data. It is implausible that a single "optimal model" can be identified since environmental variables may influence which effects are included in the model. Although this may be difficult to confirm for certain models, in general, the model that maximises genetic progress in the population should be selected for genetic evaluation.

2.4. Lactation curve models

In the context of dairy farming, the milk production characteristics of cows exhibit a specific curvilinear pattern throughout the lactation period. Knowledge about the lactation curve serves as a valuable source of information regarding the traits of milk production, which can then be used to take decisions about herd management. Recently, there has been a growing preference for utilizing test-day models in the genetic assessment of dairy cows, replacing the traditional cumulative 305-days lactation yield. Analysing the lactation curve in dairy cows has several benefits, including the ability to estimate the overall milk supply of the cows and to develop breeding and management plans that are appropriate for dairy cattle. Researchers have experimented with a variety of models to fit the lactation curve. Various models have been explored by different researchers to effectively capture the lactation curve in both indigenous and exotic cattle.

The lactation curve is a graphical illustration of the daily milk yield trend throughout the lactation period in dairy animals (Sherchand et al., 1995). The lactation curve typically comprises two phases: an initial ascending phase, demonstrating a sharp increase in milk production from parturition to attainment of peak yield, attributed to increased secretory activity of mammary epithelial cells; followed by a descending phase indicating a gradual reduction in daily milk production from peak yield to the cessation of lactation, resulting from apoptosis of mammary epithelial cells (Capuco et al., 2003). Understanding the shape of the lactation curve provides valuable insights for making informed decisions regarding culling, health monitoring, breeding strategies, and feeding management in dairy animals (Macciotta et al., 2005).

Nanda et al. (2019) reported that to estimate lactation curve characteristics with economic value, such as early yield, peak yield, time from onset to lactation peak, persistency, and total lactation yield, the lactation curve models used the function parameters. Since the lactation curve has a nonlinear tendency with time, a linear model cannot be fitted to it. Numerous mathematical models have been created to explain the flow of milk production in dairy cows throughout lactation (Bangar and Verma 2017)

A variety of mathematical models, both linear and nonlinear, have been developed to depict the pattern of lactation curves in bovine species. These mathematical models include the Inverse polynomial function (Nelder, 1966), Gamma

function (Wood, 1967), Polynomial regression function (Ali and Schaeffer, 1987), Mixed log function (Guo and Swalve, 1995), and Exponential function (Catillo et al., 2002) among others.

The lactation curve prediction depends on various environmental factors, such as the year and season of calving, parity, age of the cow, etc., the best fit of the curve is still elusive. (Catillo et al., 2002; Penchev et al., 2011)

2.4.1 Exponential decline function

Brody et al. (1923) made an early attempt to model the lactation curve, which is utilized to describe the decreasing phase of lactation as milk production typically decreases after reaching a peak shortly after calving. The reduced exponential function served as the foundation for the suggested model (Ferreira et al., 2015). This model ignores the significant rise in milk production that is frequently seen after calving and ignores the initial surge in milk production during the early stages of lactation. As a result, this model was inappropriate for illustrating the entire lactations of highly specialized cows; nonetheless, it adapted well enough to low-production crossbred animals raised in tropical climates (Cunha et al., 2010). Kolte et al. (1986) found that the exponential decline function in Sahiwal cattle justified the slightly larger (0.39) standard error of estimation and lower R^2-value (26%) and suggested that this function will be less accurate.

2.4.2 Parabolic exponential function

Sikka (1950) conducted the first attempt to focus on a model based on the lactation curve shapes of dairy animals. He presented a bell-shaped model that is commonly referred to as the "parabolic exponential model". The curve that corresponds to this function, however, differs significantly from the curve that is seen during the first stage of lactation. A moderate coefficient of determination ($R^2 = 95.00\%$) was explained by a parabolic exponential curve. Prajapati et al. (1992) observed a higher R^2-value (99.2%) for this function in Kankrej cows. However, in Hariana, Sahiwal, crossbred cows, and Sahiwal cows, respectively, lower values were observed by Singh and Bhat (1978), Yadav and Sharma (1985), Kolte et al. (1986), and Gore et al. (1996). The parabolic exponential function showed significant variation between the actual and anticipated yield to the conclusion of the lactation. As a result, the ascending, peak yield, and descending phases of lactation in Gir cows are not explained by this function.

2.4.3 Inverse polynomial function

An inverse polynomial function was described by Nelder (1966) that fits the lactation data satisfactorily but peaks earlier than the other curves. The inverse polynomial function was used by Dave (1971) and Kumar (2007) and reported 81.0% and 98.5% of R^2 values in Indian water buffaloes and Murrah buffaloes, respectively. The typical lactation curve of an Indian buffalo was found to be better described by an inverse polynomial function Kumar and Bhat (1979) than by an exponential decline function.

2.4.4 Gamma Incomplete Function

Test day data of milk yield in several dairy breeds have been fitted using the gamma type function provided by Wood (1967). Ferreira et al. (2015) reported that the lactation curve model is now improved upon over the previous old model and other models by incorporating the incomplete gamma function. A more adaptable mathematical tool that provides a better fit to both the ascending and descending stages of the lactation curve is the Gamma Incomplete Function (Wood, 1967). One of the main limitations of the Brody model (Brody et al.1923) is addressed by adding the incomplete gamma function, which gives the model the flexibility to capture the first increase in milk output following calving. Furthermore, the better fit during the falling phase implies that the model may be able to depict the drop in milk output following the peak yield more precisely. This improved mathematical framework is a useful tool for researchers and practitioners to maximize herd management and output by giving a more accurate representation of the intricate dynamics involved in dairy cattle lactation.

There have been numerous attempts to simulate the dairy cow's lactation curve. Among others, Wood (1967) work is well-known (Guo and Swalve, 1995).

The lactation curves of crossbred buffaloes were evaluated by Barbosa et al. (2007) under two production systems: PS1 (spanning from 1984 to 1998) and PS2 (from 1999 to 2002). They employed eight mathematical functions including Inverse Polynomial, Linear Hyperbolic, Incomplete Gamma, Logarithmic, Logarithmic Quadratic, Linear, Quadratic, and Jenkins & Ferrel. The incomplete gamma function emerged as the most suitable model for describing the lactation curve in both production systems, yielding high coefficient of determination values (R^2) of 95% for PS1 and 96%

for PS2. Following closely was the Jenkins and Ferrel model, with R^2 values of 0.85 for PS1 and 0.88 for PS2.

Torshizi et al. (2011) compared the fit of linear and nonlinear lactation curve models and evaluated environmental factors influencing daily milk output. The researchers utilized an extensive collection of primiparous Holstein cows' monthly test-day milk records from Tehran Province, Iran, and examined the effects of different parameters on daily milk yield. A comparison of seven mathematical models (Wood, Wilmink, Dijkstra, Rook, Cubic, Grossman, Ali & Schaeffer) with varying parameters revealed that the incomplete Gama, Dijkstra, and Grossman functions were the best models with ($R^2 = 0.999$). Because of its characteristics and computational efficiency, the Wood function was determined to be the most accurate model for estimating daily milk supply.

Bangar (2017) described the lactation curve in Gir crossbred cows using four non-linear models: the Wilmink model, the mixed log function, the quadratic model, and the gamma-type function. The first lactation produced the lowest milk yield, whereas the fourth lactation produced the most. The mixed log function and gamma-type function offered the greatest fit for the lactation curve of the first and remaining lactations, respectively. The quadratic model had the least fit in almost all lactations. When it came to describing the variability in the lactation curve in primiparous cows (first lactations) and multiparous cows (second or more lactations), the gamma-type function obtained the greatest adjusted R^2 values.

However, the Wood model has well-known drawbacks. Additionally, there have been reports of underestimating daily milk supply at and after the peak and overestimating it in the earlier portion of the curve. (Wood,1967; Dijkstra et al., 2010; Macciotta et al., 2011).

Many studies have been conducted to create linear and nonlinear lactation curve models after Wood's model. Models in the nonlinear have been produced by Wilmink (1987), while models in the linear have been developed by Ali and Schaeffer (1987) and Guo and Swalve(1995). Since the models have been shown in multiple experiments to fit data better than Wood, it is thought that they merit further investigation.

The validity of the gamma type function in predicting milk production in cows with mastitis is confirmed by greater R^2 and lower RMSE (Jingar et al., 2014).

2.4.5 Mixed log function

Guo and Swalve (1995a) developed the mixed log function originally. Guo and Swalve (1995b) used daily milk recording data from an experimental farm to examine the goodness of fit of 14 models while taking into account different comparison criteria. They stated that there were variations in the goodness of fit of lactation curve functions and that a better fit was found for second and later parities than for first lactation.

Dongre et al. (2012) examined three lactation curve models (quadratic model, gamma-type function, and mixed log function) using fortnightly test-day milk yield records of Sahiwal cows. With the greatest coefficient of determination (R^2 value) of 88.75% and the lowest root mean square error (RMSE) value of 0.08 kg, the mixed log function was determined to be the most appropriate one of all the models. The quadratic model exhibited a poor fit to the observed lactation curve, with the lowest R^2 value of 56.93% and the highest RMSE value of 0.15 kg. Although the gamma-type function produced a lower peak yield than the mixed log function, it suited the actual lactation curve more closely. It is advised for cows with low yields. According to Catillo et al. (2002), the mixed log model worked well for calculating the lactation curves of Italian water buffalo milk production features. Using a mixed log function, Cilek and Keskin (2008) observed an R^2-value of 92.7% in Simmental cows.

2.4.6 Multiphasic logistic function

According to Grossman and Koops (1988), the diphasic logistic function (n=2) has advantages over the gamma function since it has meaningful functions with easily interpretable parameters and minimum residuals. According to De Boer et al. (1989), the triphasic logistic function (n=3) was favoured over the diphasic function due to its smaller and less auto-corrected residuals. According to Kocak and Ekiz (2008), the model's R^2 value for the first lactation in Holstein cattle was 59.3%. For the first lactation of Karan Fries cattle, Banu (2010) observed root mean square error of 43%, 46%, and 17% of monophasic, diphasic, and triphasic functions, respectively.

2.4.7 Polynomial regression function

In HF cows, it was discovered that the polynomial regression function was slightly more accurate than the wood's function and that both of them were better than the inverse quadratic polynomial function (Ali and Schaeffer, 1987).

Drawing from the lactation records of primiparous lactating Murrah buffaloes, Sahoo et al. (2015) explored the prediction of first lactation 305DMY using bi-monthly test-day milk yields through the application of four mathematical models: the Incomplete Gamma Function, Exponential Function, Mixed Log Function, and Polynomial Regression Function. Among these models, the polynomial regression function (PRF) developed by Ali and Schaeffer exhibited the best fit, demonstrating the highest coefficient of determination (R^2) value of 0.998 and the lowest Root Mean Squared Error (RMSE) value of 0.03. Consequently, the researchers concluded that PRF proved to be the most effective function for predicting 305DMY in Murrah buffaloes. This finding aligns with their previous study on milk yields based on fortnightly test days in Murrah buffaloes (Sahoo et al., 2014), where they reported an R^2 value of 0.997 and an RMSE value of 0.02.

Using the polynomial regression function on Italian water buffaloes Kumar (2007) reported R^2-values of 96.7%. According to Kumar (2007), the most accurate model for forecasting the milk production on test days, both weekly and monthly, in the first three and pooled lactations of Murrah buffaloes was the polynomial regression function.

Singh et al. (2015) concluded that the polynomial regression function followed by the Exponential function method was found to be the most appropriate fitted method for predicting the monthly test day milk yield data in Murrah buffalo. It is also reaffirmed by many other researchers, indicating that the PRF model was the best fit by Macciotta et al. (2005) in Italian Simmental cows with $R^2 > 0.75$, Torshizi et al. (2011) by Iranian Holstein cows with $R^2 = 0.999$, RMSE = 0.001 and by Kumar (2007) in Indian Murrah buffaloes with $R^2 = 99.7\%$.

2.4.8 Exponential function

The exponential function was initially created by Wilmink (1987). Though it did not fit well with individual lactation in HF cows, this function was shown to be the best three-parameter model for predicting herd mean yield (Olori et al., 1999). According to Silvestre et al. (2006), there was a 0.93 correlation between the estimated milk yield and daily milk yield, indicating the degree of association between the real and calculated values.

Table 2.2: Several widely used empirical models for fitting the lactation curve in dairy animals

Breed	R^2- value (%)	References
Parabolic exponential function		
Hariana	93.6	Singh and Bhat (1978)
Rathi	74.0	Gahlot et al. (1988)
Kankrej	99.2	Prajapati et al. (1992)
Gir	68.1	Pundir and Kaushik (1993)
Sahiwal	80.5	Dongre (2012)
Inverse polynomial function		
Indian water buffalo	81.0	Dave (1971)
Hariana	99.9	Singh and Bhat (1978)
Crossbreed	90.0	Yadav and Sharma (1985)
Rathi	99.7	Gahlot et al. (1988)
Gir	90.3	Pundir and Kaushik (1993)
HF	97.9	Olori et al. (1999)
Murrah buffalo	98.5	Kumar (2007)
Sahiwal	99.8	Dongre (2012)
Gamma Incomplete Function		
Indian buffalo	96.0	Kumar and Bhat (1979)
Murrah buffalo	89.0	Cheema and Basu (1983)
Italian water buffalo	93.7	Dimauro et al. (2005)
Egyptian buffalo	96.0	Aziz et al. (2006)
Mediterranean Italian buffaloes	95.0	Coletta et al. (2007)
Murrah buffalo	96.2	Katneni (2007)

Breed	R^2- value (%)	References
Holstein	59.5	Kocak and Ekiz (2008)
Karan Fries	87.9	Banu (2010)
Jersey	91.5 to 93.3	Cankaya et al. (2011)
Sahiwal	95.6	Dongre (2012)
Mixed log function		
Holstein Friesian	96.4	Olori et al. (1999)
Irish cattle	64.0	Quinn et al. (2005)
Murrah buffalo	96.7	Kumar (2007)
Simmental cow	92.7	Cilek and Keskin (2008)
Sahiwal	97.9	Dongre (2012)
Exponential function		
Hariana	95.0	Yadav et al. (1977)
Jersey	55.0	Aguirre and Boschini (1992)
Gir	40.1	Pundir and Kaushik (1993)
Italian water buffalo	94.4	Dimauro et al. (2005)
Murrah buffalo	99.6	Kumar (2007)
Sahiwal	67.7	Dongre (2012)
Polynomial Regression Function		
Italian water buffalo	96.7	Dimauro et al. (2005)
Murrah buffalo	99.7	Kumar (2007)
Simmental	95.1	Cilek and Keskin (2008)
Holstein	98.5	Sharifi et al. (2009)
Karan Fries	99.5	Banu (2010)

According to Rowlands et al. (1982), data from high-producing dairy cows at the peak of their milk production do not consistently fit well into Wood's curve. Regression models for milk yields on the day of lactation (quadratic and linear) and on the log of 305-d divided by the day of lactation (quadratic and linear) were shown to be superior to the Gamma function (Ali & Schaeffer, 1987). The Dijkstra function outperformed the Wood, Wilmink, and Rook functions in defining lactation curves in dairy cattle maintained on farms in central Mexico, according to a study conducted on the subjects (Val- Arreola et al., 2004).

Table 2.3: Different conventional lactation curve models

Function	Author	Models
$y = ae^{-bt}$	Brody et al. (1923)	Exponential Model
$y_t = at^b e^{ct}$	Wood (1967)	Gamma Incomplete Function
$y_t = ae^{-bt} - ae^{ct}$	Cobby and Le Du (1978)	Cobby and Le Du Model
$y_t = a + be^{kt} + ct$	Wilmink (1987)	Wilmink Model
$yt = a6\, t^{b6}\, e^{-c6t}\,(1 + u\sin(x) + v\cos(x))$	Grossman et al., 1986	Grossman model
$Y_t = a + bx\sqrt{t} + cx\log(t)$	Guo and Swalve 1995	Mixed log function
$Y = t\,(a + bt + ct^2)^{-1}$	Nelder 1966	Inverse polynomial model
$Yt = a + bx + cx^2 + d\log(1/x) + f\log(1/x)^2$	Ali and Schaeffer (1987)	Polynomial Regression Function
$Y_t = a\exp(-bt + ct^2)$	Sikka (1950)	Parabolic Exponential Model

For cows with incomplete or ongoing lactations, Vasconcelos et al. (2004) and Macciotta et al. (2002) assessed how well an autoregressive multiple-lactation test day (AMLTD) model predicted the test day yields of milk, fat, and protein that were missing. In comparison to the test interval approach, the ALMTD procedure demonstrated strong predictive power and produced fewer biased estimates of lactation yield, making it a valuable tool for estimating cumulative lactation yields in dairy cattle assessment systems. Graesbøll et al. (2016), on the other hand, used lactation curves with fewer parameters to develop a reliable prediction model for milk yield at the cow level, which is helpful when there is sparse data.

According to Schaeffer and Dekkers (1994), one way to account for differences in the lactation curve's shape is to apply a random regression within the cow. Bilal and Khan (2009) reported that, for predicting the test day model, the random regression method is superlative in comparison to the fixed regression model and repeatability test day model. The random regression model is more efficient than the traditional model

in Gir cattle (Pereira et al., 2019) and gives higher heritability than the test day model and multiple trait models in Sahiwal cows (Bilal et al., 2009).

For routine genetic evaluation of production parameters in dairy cattle, the Random Regression Model (RRM) methodology is widely utilized. The RRM has the following benefits: it allows for the analysis of the environmental factors unique to each TD; it permits each animal's lactation curve to have a unique shape; it provides for the calculation of EBV for persistency or partial lactation yields; and it allows animals to be assessed based only on a single TD yield. Jamrozik et al. (1996). RRM's primary concept is to fit average lactation curves for a subpopulation and animal-specific curves that explain average curve variances. Fixed and random regressions can be fitted with a variety of functions. In the world of dairy farming, lactation yield is the amount of milk produced by a dairy cow over a given period and it is extremely important. It serves as a pivotal parameter that demands daily recording, as its accuracy and consistency are essential for effective selection and management within the dairy industry. In essence, lactation yield, when recorded on a daily basis, serves as an invaluable tool that empowers dairy professionals to optimize their operations, resulting in healthier, more productive herds. However, recording of daily milk yield data poses a number of serious difficulties. The amount of milk produced fluctuates not only from day-to-day measures but also depends on a wide range of effects, such as the health of the cow, the environment it lives in, and the nourishment it consumes. To guarantee an accurate representation of milk output, this intricacy necessitates frequent and exact measurements.

For a variety of prediction tasks in the dairying industry, conventional regression models have been widely employed (Sundaresan et al., 1954; Puri & Sharma, 1965; Schaeffer et al., 1977; Jain & Taneja, 1984; Gandhi & Gurnani, 1988; & Geetha et al., 2006). Typically, these models rely on specific presumptions that are necessary for them to function correctly. Assumptions are, nevertheless, occasionally broken in real-world scenarios. The Animal Improvement Programme requires the development of a predictive model to anticipate the economic features that determine whether animals are accepted or rejected for inclusion in the herd's next generation (Sharma, 2011).

When it comes to fitting the lactation curve, the Wood's model, like lactation models uses a mix of linear and non-linear regression techniques to estimate its parameters and calculate characteristics of the lactation curve (Macciotta et al., 2011).

However mathematical models do have their limitations. Firstly they often restrict the number of input data points that can be considered, meaning a few inputs are included out of the many that may be available. Secondly these models typically assume a relationship function for the entire dataset, oversimplifying the potential complexity of relationships. Thirdly they rely on assumptions imposed by their theories such as normality, linearity and data independence. While these assumptions are crucial for models to work, but they may not hold true in real world situations and compromising their effectiveness (Kominakis et al., 2002).

Modeler's are now focusing on linear functions like polynomials or splines due to the increasing availability of individual lactation records and genetic evaluations based on test day records. With more animal records being collected and advancements in farm management software, there is a growing need to predict individual variations, from the mean lactation curve over time (Macciotta et al., 2011).

Non-linear regression is known for its enhanced flexibility and accuracy, in lactation curves surpassing the capabilities of regression. However it should be noted that it may be more vulnerable to the influence of noise, outliers or multicollinearity, in the data when compared to machine learning algorithms.

2.5 Machine learning models

In the field of animal breeding, effectively managing huge amounts of production related data is crucial. The transition from statistical models to machine learning has proven to be highly advantageous in this regard (Figure 2.1). Unlike statistical models, machine learning models are particularly skilled at processing and interpreting datasets. This offers breeders a way to predict outcomes and automate processes. By utilizing algorithms these machine learning models can identify patterns within the data leading to more accurate decision making and valuable insights. This shift not streamlines processes but also empowers breeders to optimize their strategies based on real time information. Ultimately it enhances productivity and overall success in animal breeding programs.
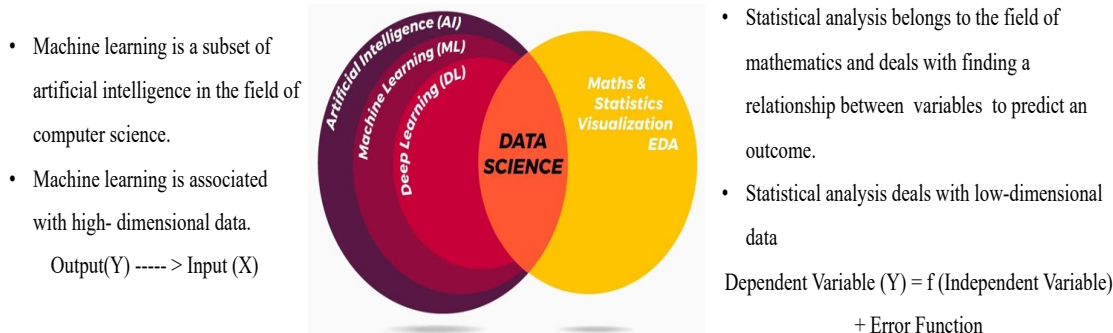
- Machine learning is a subset of artificial intelligence in the field of computer science.
- Machine learning is associated with high- dimensional data.

    Output(Y) ----- > Input (X)

- Statistical analysis belongs to the field of mathematics and deals with finding a relationship between variables to predict an outcome.
- Statistical analysis deals with low-dimensional data

Dependent Variable (Y) = f (Independent Variable) + Error Function

Figure 2.1: Relationship between Machine learning and Statistical analysis

Machine learning systems are a subset of intelligence that draw inspiration from the brain and its neural networks. Similar to human learning, machine learning algorithms can perform tasks such, as classification, pattern recognition, optimization and making projections based on acquired knowledge gained through repeated experiences (Tadeusiewicz 1993; McQueen et al., 1995; Kominakis et al., 2002; Michalski et al., 2013).

Machine learning involves generating models capable of making predictions based on acquired knowledge. These algorithms analyse data records that contain variables. They use this information to identify a pattern that describes an outcome. This process of acquiring knowledge is commonly referred to as "training ". Afterwards the models are subjected to "testing" or "validation" using data that was not part of the training set (Shahinfar et al., 2012; Hempstalk et al., 2015).

Although machine learning has achieved success in various fields, there has been limited research on its applications in case of animal breeding (Kominakis et al., 2002; Michalski et al., 2013; McQueen et al., 1995; Kotsiantis et al., 2007).

In the dairy industry, to improve the efficiency of the milk yield prediction by inferring the expected lactation curve, a reliable reference of milk yield records is required to be compared with the actual milk yield. However the available data in this sector is often unstructured and poorly integrated making traditional modelling techniques ineffective for analysis. To overcome this challenge and model the lactation curve accurately we can employ Machine Learning (ML) an flexible method (Figure 2.2) that allows us to analyse large datasets effectively (Lacroix et al., 1995).
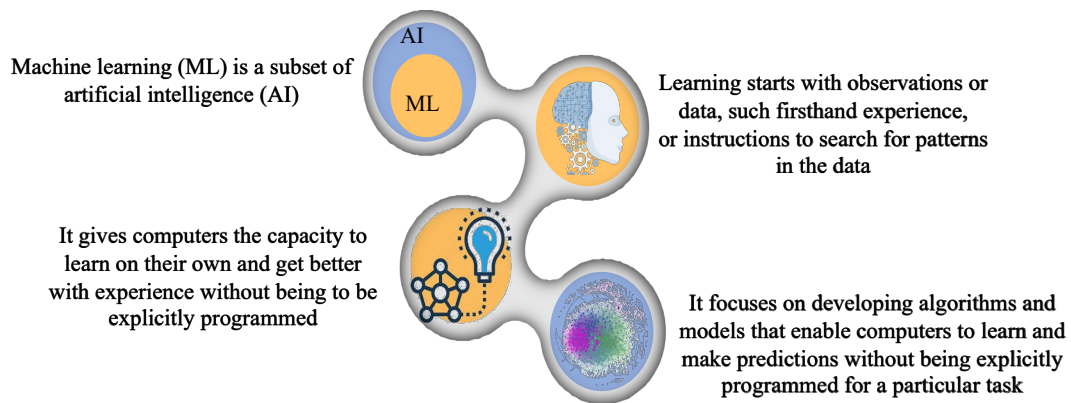
Machine learning (ML) is a subset of artificial intelligence (AI)

AI
ML

Learning starts with observations or data, such firsthand experience, or instructions to search for patterns in the data

It gives computers the capacity to learn on their own and get better with experience without being to be explicitly programmed

It focuses on developing algorithms and models that enable computers to learn and make predictions without being explicitly programmed for a particular task

Figure 2.2: Components of Machine learning

2.5.1 Artificial Neural Network

An artificial neural network (ANN) is a computational model inspired by biological structure of the human brain systems, comprising numerous individual units known as artificial neurons. These neurons are interconnected with coefficients (weights), forming the neural structure referred to as processing elements (PE), these neurons serve to process information. Each PE involves weighted inputs, a transfer function, and produces one output. Essentially, a PE can be described as an equation that balances inputs and outputs. ANNs are alternatively termed connectionist models, highlighting the significance of connection weights in their functioning.

The effectiveness of neural computations arises from the interconnection of neurons within the network (Kustrin & Beresford 2000). Unlike traditional programming, ANNs acquire knowledge through experience and learning rather than explicit programming. Artificial Neural Networks (ANNs) can process any type of input. Optimizing network design, training cycles, input combinations, learning parameter values, and data pre-treatment, however, may enhance their performance.

Recent years have seen investigations on artificial neural networks in animal husbandry. They have been applied to illness detection, animal breeding, quality assessment, and yield characteristic and classification prediction. (Atıl and Akıllı 2016). The link between input and output variables in linear and nonlinear systems can be found using this highly effective mathematical modelling technique. But the results of the ANN analysis often have low error rates, based on traditional statistical approaches. (Takma et al., 2012; Akıllı and Atıl, 2014). Different regression models are commonly employed in milk yield analyses to predict future yields. However,

artificial neural networks (ANN) offer a distinct approach compared to traditional regression methods or time-series analyses. Unlike conventional methods that involve constructing an algorithm and implementing it as a computer program, ANN tackles specific problems through learning. This learning process is based on typical inputs and their corresponding desired responses, providing a more dynamic and adaptable approach to yield prediction in comparison to more traditional techniques (Grzesiak et al., 2003).

While traditional lactation models generally produce good results on average, there is need for improvement when it comes to estimating a cow's milk production in the lactation process. Artificial neural networks (ANNs) known for their prowess in pattern recognition offer an alternative to conventional models (Lacroix et al., 1995). The prediction of dairy yield poses a challenge that has benefited from the application of statistical techniques. Recently, ANNs have emerged as a substitute method for forecasting milk yield (Njubi et al., 2009).

In various countries, the genetic assessment of dairy cattle is based on analyzing the milk yield over a 305 day lactation period. To estimate milk yield several regression models are commonly used to project production levels. However artificial neural networks (ANNs) have gained recognition as an alternative method to regression techniques and time series analysis. Grzesiak et al. (2003) demonstrated that using ANN and Multiple linear regression (MLR) yielded no significant difference (P >0.05) between projected values from observed values, indicating that ANN was indeed a model suitable for predicting 305 day milk production.

According to a study by Lacroix et al. (1995) artificial neural networks (ANNs) have demonstrated the ability to predict milk production in cows effectively and earlier. This advancement is particularly important, during the stages of lactation, where accurately forecasting a 305 day milk supply can greatly influence the selection of bull dams.

In a study conducted in Kirsehir, Turkey by Gorgulu (2012) explored the use of networks (ANNs) to predict the 305 day milk yield of 2,640 Brown Swiss cattle. The study took into account factors such as age, number of lactation and season of calving. Comparing ANNs with the multiple linear regression method it was found that ANNs provided more accurate predictions than MLR. The mean predicted values calculated

by ANNs closely matched the values of milk production with correlation coefficients (R) ranging from 0.74 to 0.82 for the artificial neural networks. However there was a disparity between the mean values and the expected mean values determined by MLR. The design of the ANN utilized MATLAB 7.0. Followed a three layer architecture consisting of input, hidden and output layers with 4, 8 and 1 neurons. To. Validate the ANN model, a training set, verification set and test set were used. To assess the effectiveness of the ANN models performance metrics such as root mean square error (RMSE) were employed.

Predicting milk yield in dairy animals is a complex task due to the interaction between genetic and environmental factors. While mathematical models have been utilized for predicting lactation yields they have limitations in terms of input data quantity and assumptions related to linearity and data independence. Artificial neural networks (ANNs) are proposed as a tool for milk yield prediction as they can learn from historical data without relying on predefined models or identifying all variables in advance. ANNs have proven successful in dairy cattle breeding by predicting milk production and detecting diseases. Kominakis et al. (2002) aimed to explore the potential of ANNs for predicting test day and total milk yield in dairy sheep based on test day records at the beginning of lactation.

Njubi et al. (2009) demonstrated that ANN models can be used as an alternative to monthly test day, for Kenyan Holstein cows specifically in predicting first lactation 305 day milk yields (FL305DMY). Despite outperforming multiple linear regressions in terms of R^2 and root mean square error there was a significant difference observed (p>0.05). When estimating the FL305DMY from TD yield, the prediction equation with four variables as first, second, third, and fourth TD milk yield provided a satisfactory level of accuracy (79.0%).

The study by Kumar et al. (2019) used monthly test day milk records from Murrah buffalo; the accuracy of predicting FL305DMY using ANN was almost identical to that of linear regression. While the multiple linear regression attained 76.9% accuracy in predicting FL305DMY in Murrah buffalo, the best ANN algorithm achieved 76.8% prediction accuracy for the model.

The accuracy of milk production predictions can be significantly influenced by processing techniques, a well-designed network model and careful selection of

variables. This has been demonstrated by studies conducted by Lacroix et al. (1995) and Salehi et al. (1998b) who utilized networks in their milk yield predictions.

In the later part, various artificial neural networks (ANN) have been introduced over time. A significant breakthrough was made by Lacroix et al. (1995) who successfully developed the multilayer perceptron model (MLP) that could forecast the 305 day yield based on 16 variables. Subsequent improvements included advanced data processing techniques, by Lacroix et al. (1997) and the utilization of multiple networks each specializing in specific predictions as demonstrated by Salehi et al. (1998). MLPs have also been used to predict aspects of milk yield, such as the 305 day milk yield (Grzesiak et al., 2003a; Gorgulu, 2012) the lactations 305 day milk yield (Sharma et al., 2006, 2007; Njubi et al., 2010) daily milk yield (Grzesiak et al., 2006; Torres et al., 2005) and even overall dairy herd production (Murphy et al.,2014; and Sanzogni & Kerr, 2001). These advancements highlight the adaptability and efficiency of MLPs in utilizing networks, for contexts in predicting milk yield.

2.5.2 Machine learning linear regression models

Machine learning linear regression models (MLR) stands out as one of the most straightforward and widely employed machine learning algorithms. It operates as a mathematical technique for conducting predictive analysis, enabling the projection of continuous, real, or mathematical variables. (Maulud & Abdulazeez, 2020, Hempstalk et al., 2015). Linear regression serves as a mathematical test employed to assess and quantify the relationship between the variables under consideration (Abdulqader et al., 2020; Akgün & Öğüdücü, 2015; Dehghan et al., 2015).

Linear regression is frequently utilized in mathematical research methodologies, allowing the measurement and comparison of predicted effects against various input variables. It serves as a technique for evaluating and modeling data, establishing linear connections between dependent and independent variables. This method facilitates the modeling of relationships, discerning connections between dependent and independent variables based on the analysis and learning derived from current training outcomes.(Lim 2019).

Linear regression techniques demonstrate limited adaptability and suboptimal predictive accuracy when the association between input and output variables cannot be adequately modelled by linear functions.

2.5.3 Random Forest

Random Forest (RF), a machine learning algorithm introduced by Breiman in 2001 has gained popularity in fields due, to its effectiveness. Its successful application can be observed in animal science studies conducted by Shahinfar et al. (2014), Brieuc et al. (2018), Gianola et al. (2011), Guevara et al. (2023).

The random forest algorithm functions, as either a regression or classification method utilizing decision trees that strategically divide the data to minimize variation. Each tree is trained on a subset of the data and during the tree building process a random subset of variables is selected at each split (Breiman 2001). Although individual trees may not be classifiers individually they are combined into an ensemble called a forest. This ensemble, formed by merging the predictions of trees serves as the model. The final predictions generated by a forest are based on the outcomes of numerous individual tree predictions (Van der Heide et al., 2019).

2.5.4 eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGBoost) is a widely-used machine learning algorithm for regression tasks, leveraging ensemble learning principles to build a series of decision trees. Sequentially constructed, each tree corrects errors made by the existing ensemble, minimizing the difference between predicted and actual values through a gradient boosting approach. To enhance model generalization, XGBoost incorporates regularization techniques and tree pruning, preventing overfitting by penalizing complex models and controlling tree depth. The algorithm also provides insights into feature importance, aiding in feature selection and model interpretation. Effective hyperparameter tuning is crucial for optimizing XGBoost's performance, with parameters governing learning rates, tree depth, and regularization terms. Its adaptability and robustness make XGBoost a popular choice for regression challenges across diverse domains.

2.5.5 Sequential Minimal optimization regression (SMOreg)

Sequential Minimal Optimization Regression (SMOreg) is a machine learning technique commonly used for regression purposes (Shevade et al., 2000). It is built upon the Sequential Minimal Optimization (SMO) approach originally designed for training Support Vector Machines (SVMs). SMOreg functions by selecting pairs of data points. Adjusting the model parameters to reduce the difference between predicted and actual values. This method is effective for regression tasks that involve forecasting

output variables (Guevara et al., 2023). SMOreg is recognized for its efficiency in managing datasets and its accuracy in making predictions. It proves beneficial in situations with input features requiring high dimensional regression analysis. SMOreg serves as an asset, within the realm of machine learning for addressing regression challenges.

2.6 Sire evaluation

Enhancing cow productivity involves employing high-quality sires, necessitating a genetic assessment of sires for milk production. The success of genetic enhancement depends on accurate selection, selection intensity, genetic trait variability, and the generation interval. Breeders seek effective methodologies to expedite genetic improvement by selecting animals with superior breeding values for economically significant traits (Girimal et al., 2022). Breeding values can be determined through individual, family, or combined phenotypic data. It is possible to assess sires across one or several herds. The statistical equations depict approaches for sire evaluation as either non-linear or linear. As developments are achieved in sire evaluation based on data structures, breeding methodologies, and selection processes, a multitude of ways can be implemented.

An individual's genetic worth can be determined using a wide range of sire evaluation techniques, including the Simple Daughter Average (D), contemporary comparison method (CC), Least Squares Method (LSM), Simple Regressed Least Squares (SRLS), Best Linear Unbiased Prediction (BLUP), Derivative-Free Restricted Maximum Likelihood (DFREML) and WOMBAT for both single and multiple trait models.

The assessment of different sire evaluation methods involved the consideration of various criteria such as within-sire variance, error variance, coefficient of determination (%), coefficient of variation (%), and rank correlations. The method demonstrating the lowest error variance was deemed the most effective. Greater coefficients of determination (R^2-value) obtained from model fitting indicated higher accuracy. The stability of a sire evaluation method was determined by its ability to maintain the coefficient of variation (CV%) close to the unadjusted data's CV%. A higher rank correlation, approaching unity, among sires evaluated using different methods indicated a greater similarity in the ranking obtained from diverse evaluation approaches (Singh 2016).

2.6.1 Simple Daughter Average

The evaluation of sires was conducted through the Simple Daughter's Average (Robertson and Rendel, 1954). This approach involves straightforward computation to determine the breeding value of sires, enabling the prediction of the performance of future daughters born to cows with similar genetic potential as the evaluated sire. Edward (1932) assessed the breeding value of sires by calculating the average index of their daughters, expressed as

$$I = D$$

Where,

I = the sire index

D = the average of daughters of the sire.

2.6.2 Least squares method

Robertson & Rendel (1954) first introduced the least squares technique for assessing and evaluating sires. The fundamental idea of the least-squares principle is to reduce error variance by making adjustments for diverse non-genetic factors within the data.

The LSM principle relies on minimizing the square of the disparity between observed and predicted values of the dependent variables, striving for the least possible difference, ideally approaching zero. (Talokar et al., 2023). The process involves sequentially estimating the actual effects and predicting the breeding values of sires within a population based on statistical data.

Cunningham (1965) outlined a technique for acquiring weighted least squares estimates of sires using non-orthogonal data from progeny test records, particularly in cases where artificial insemination (AI) was employed. He noted the feasibility of categorizing sires into distinct groups at a much earlier stage in their development, even before the proofs were finalized.

Jain and Malhotra (1971) discovered that the index $I = A + 1/2 h^2 q (D - CD)$ proved to be the most effective for evaluating dairy sires when dam's records were not accessible. Alternatively, the index $I = A + 1/2 h^2 Q (D - CD) - b (M - CM)$ was identified as the best when considering the relative merit of eleven methods for estimating breeding values of dairy sires on a single herd basis, particularly when dam's records were available.

Tajane and Rai (1990) applied the least squares method to estimate the breeding value of 29 Holstein-Friesian sires and 8 Sahiwal sires, utilizing data from 1257 (Holstein-Friesian X Sahiwal) and 519 (Sahiwal X Holstein-Friesian) daughters.

Meanwhile, Gandhi and Gurnani (1991) determined the breeding value of Sahiwal sires by employing the least squares technique and focusing on the first lactation with a 305-days or less milk yield from their daughters, which were maintained at five different farms.

Singh et al. (1992), Raheja (1992) employed the least squares method to estimate the breeding values of Hariana bulls and Sahiwal sires respectively. Parekh et al. (1994) conducted an evaluation of Friesian, Jersey, and Brown Swiss sires, utilizing the least squares method for estimation.

Raheja (1992) indicated that the accuracy of the least squares (LSM) method surpassed that of the best linear unbiased prediction (BLUP) method.

Banik (2004) employed the Least Squares method, along with other techniques such as the Contemporary Comparison method, SRLS, BLUP, and DFREML, to evaluate Sahiwal sires. He reported highly significant rank correlations of LSQ with the Contemporary Comparison method (0.91), SRLS (0.98), BLUP (0.85), and DFREML (0.96). These results suggested that the ranking of sires remained consistent across these methods, indicating minimal alteration in the evaluation outcomes.

The least squares method for determining the genetic worth of sires is grounded in the fundamental principle of minimizing the error variance after accounting for various non-genetic or environmental factors.

2.6.3 Simple Regressed Least Squares

Dhaka and Raheja (2000) employed DLS, RLS, and BLUP procedures for the assessment of Sahiwal bulls. They noted that estimates of breeding value obtained from RLS exhibited a perfect normal distribution, followed by BLUP. Additionally, RLS demonstrated the minimum standard error, followed by BLUP and OLS methods. However, considering the computational costs and relative computational difficulties, they suggested that BLUP stands out as the most suitable method.

Similarly numerous researchers in India have employed this method for estimating the breeding value of sires. Notable studies by Gandhi and Gurnani (1991),

Raheja (1992), Pundir and Raheja (1997), Deulkar and Kothekar (1999), Pundir et al. (2004), Banik and Gandhi (2010), Bajetha (2006), Kumar (2007), Kathiravan (2009), and Raja (2010) utilized this approach to evaluate Sahiwal sires. Similarly, Tajane and Rai (1990), Singh et al. (1992), Parekh et al. (1994), Jain (1996), and Mukherjee (2005) applied this method to assess sires of other indigenous breeds.

2.6.4 Best Linear Unbiased Prediction

Henderson (1949; 1973) introduced the Best Linear Unbiased Prediction (BLUP) method, which has proven to be a superior approach (Henderson, 1973; Falconer, 1989) especially when employing the Individual Animal Model (IAM) for analysis. The BLUP analysis model incorporates fixed effects and considers the relatedness between animals. Consequently, breeding values estimated through this method are less prone to bias (Falconer, 1989; Henderson, 1973).

Henderson (1973) introduced the concept of the Best Linear Unbiased Prediction (BLUP) method for sire evaluation within mixed model equations. This approach amalgamates aspects of both least squares and selection index techniques, establishing itself as a potent and adaptable method (Henderson, 1974). Known for its unbiased nature, the BLUP method provides predicted values that closely align with the expected values of sires while minimizing error variance. Additionally, the method is easily modifiable to accommodate changes in conditions.

A study on Hariana cows conducted by Dahiya et al. (2005), utilized auxiliary traits on daughters along with first lactation milk yield (FLMY) to evaluate sires for individual traits using various methods, including Simple Daughter Average, Contemporary Comparison, Least Squares, Sire Restricted Least Squares (SRLS), and Best Linear Unbiased Prediction (BLUP). Their findings revealed that BLUP emerged as the most efficient method for sire evaluation, exhibiting the lowest error variance and highest relative efficiency compared to other methods.

Singh et al. (1992) conducted a comparison of the BLUP, LSQ, SRLS, and CC methods to rank Hariana sires based on part lactation milk yield. The study concluded that the BLUP method was deemed more appropriate than the others, attributed to its robust theoretical foundations and lower prediction error variance.

2.6.5 Derivative-Free Restricted Maximum Likelihood

An assessment of Frieswal sires for milk yield by Mukherjee (2005), employing "least-squares", "best linear unbiased prediction" (BLUP), and "derivative-free restricted maximum likelihood" (DFREML) methods revealed that the DFREML method demonstrated the lowest error variance, establishing itself as the most efficient among the three sire evaluation methods.

Similarly, Espinosa et al. (2001) utilized 2618 milk production records of Holstein dairy herd and applied a derivative-free algorithm within a REML framework to estimate the variance components. The study concluded that the derived variance components were reliable for predicting the breeding values of the animals.

CHAPTER III

MATERIALS AND METHODS

3.1 Sources of data

The data for the current research were systematically collected from records maintained at the Directorate of Livestock Farms, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana. This comprehensive dataset spans a period from 2014 to 2022, thus encompassing a total duration of 9 years. The dataset includes pivotal information pertinent to the investigation, notably including data on production and pedigree. The compilation process involved extraction of data from diverse registers, specifically the calving registers, milk yield registers, production registers, dry cow registers, and pedigree registers. Particular attention was directed towards the parameter of milk yield, with detailed records documenting the daily milk yield of individual animals. The milk yield records encompassed in the dataset provide a nuanced understanding of the animals' productivity, offering insights into variations across different species. Additionally, comprehensive data such as animal identification particulars, parity, lactation length, pedigree, and, notably, morning and evening milk yield for each individual of Murrah and Nili Ravi buffaloes and crossbred cattle, were methodically included in the analysis, augmenting the richness and depth of the dataset for academic scrutiny and facilitating a comprehensive exploration of the factors influencing daily milk production. It is noteworthy that records exclusively encompass animals that completed their lactations during their tenure within the herd, thereby contributing to the precision and relevance of the research findings.

3.2 Location of the farm and environment

The Livestock Farm, at Guru Angad Dev Veterinary and Animal Sciences University is located in Ludhiana, Punjab. It is located at an elevation of 244 meters above sea level. Its geographical coordinates are approximately 30.90° N latitude and 75.80° E longitude. The farm is situated in the Indo-Gangetic Alluvial plains, which has a sub-tropical climate with significant temperature fluctuations. During Winter minimum temperature drops to around 2°C while summer temperatures can reach up to 45°C. The farm locale receives rainfall ranging from 760 to 960 mm with the majority occurring in July and August. Relative humidity on the farm locale varies between 40% and 70%. These environmental factors expose the livestock on the farm to diverse

weather patterns, including extreme ones. As a result, it is crucial to implement livestock management practices.

3.3 Management of herd

The herd management for the animals under investigation involved housing them in groups within a loose housing system, situated in an area with a partially covered shelter at the Directorate of Livestock Farms. The housing arrangement included separate facilities for various categories of animals, such as calves, heifers, milking, dry, advanced pregnant females, and breeding bulls. Within this setup, the animals enjoyed unrestricted access to both feed and water. A total mixed ration (TMR) was provided, comprising a combination of greens, dry fodder, concentrate, mineral mixture, and other feed additives. The formulation of the TMR was tailored to meet the specific nutritional requirements corresponding to the different stages of the animals, ensuring their well-being and optimal development.

Daily milk yield measurements were diligently recorded during individual milking sessions for each animal. Reproductive activities were managed through artificial insemination, a method employed to control and enhance the breeding processes within the herd. This comprehensive approach to herd management encompassed provisions for nutrition, shelter, and specialized care based on the specific life stages and reproductive needs of the animals, contributing to the overall health and productivity of the herd.

3.4 Milk yield recording system

The milk yield recording system at the Directorate of Livestock Farms of Guru Angad Dev Veterinary and Animal Sciences University serves as a vital tool for comprehensive herd management, focusing on the accurate monitoring of daily milk yield from individual animals. The calves undergo the weaning process immediately upon calving. Subsequently, the milk recordings for these cattle and buffaloes commenced on the first-day post-calving and continued until the drying of an animal. Throughout this period, the animals were subjected to twice-daily milking sessions. This system uses modern technology, including automated milking machines for efficient and precise data collection. The choice between machine and hand milking methods was determined by the animals' individual milk yield and their respective adaptation levels to machine milking. Each animal is uniquely identified, enabling

personalized tracking and facilitating the correlation of milk production data with nutritional and health parameters. The integrated approach provides valuable insights into herd productivity, supports optimized feeding programs, and aids in the identification of superior genetic traits for targeted breeding programs.

3.5 Structure of data

The data structure for the study encompasses two breeds of buffaloes: Murrah and Nili Ravi, along with the crossbred cattle herd. Within each breed, the dataset is organized to include information on several key variables. These variables comprise the number of animals in each breed, offering insights into the breed-specific population. Furthermore, the dataset incorporates crucial parameters related to milk production, such as total yield, date of calving, lactation length, and daily milk yield measured both in the morning and evening. This comprehensive structure allows for a thorough analysis of breed-specific characteristics, and milk production dynamics, reproductive patterns, offering a holistic view of the populations.

3.6 Collection of data

In the present study, a comprehensive set of observations encompassing three key categories -identification and pedigree records, important events, and milk yield-related data collected. These data gathered from 2014 to 2022, were sourced from the farm, encompassing Murrah and Nili Ravi buffaloes, as well as crossbred cattle. The collected data are utilized to assess individual animal 305DMY. These observations have served as invaluable tools, enabling informed decisions aimed at enhancing performance and sustainability within the herd.

- Identification and Pedigree records
  - Animal ID
  - Species
  - Breed
  - Sire and Dam
- Important Events
  - Date of Birth
  - Date of Service
  - Date of successful service
  - Date of Calving
- Milk Yield related records

- Daily milk yield

- Peak yield

- Days to attain peak yield

- 305-days (or less) milk yield

- Lactation length

3.7 Classification of data for non-genetic factors

3.7.1 Season of calving

Season of calving (SOC) is recognized as a primary environmental element influencing the productivity of the dairy animals. Throughout the year, climatic conditions exhibit substantial variations, with more pronounced differences observed between seasons rather than within them. Therefore, the annual cycle is categorized into five distinct seasons based on Indian Meteorological Department - Punjab Table 3.1.

Table 3.1: Classification of season

Season	Month of calving
Winter	1st December to 31st January
Spring	1st February to 15th April
Summer	16th April to 30th June
Rainy	1st July to 15th September
Autumn	16th September to 30th November

3.7.2 Period of calving

The study spanned a total of nine years, from 2014 to 2022, which were segmented into three periods based on the year of calving, with intervals of three years each, as outlined in Table 3.2.

Table 3.2: Classification of period

Period	Year of calving
I	2014-2016
II	2017- 2019
III	2020-2023

3.7.3 Parity

The table categorizes dairy animals based on their lactation number, by classifying animals into different parity groups, ranging from first parity to fifth and subsequent parities, Table 3.3 offers a concise overview of the stage of lactation cycles.

Table 3.3: Classification of parity

Lactation number	Parity
First parity	1
Second parity	2
Third & Fourth parity	3
Fifth and more	4

3.8 Data streamlining

Data streamlining involves organizing and optimizing the collected data to enhance efficiency in subsequent analyses. This process may include removing redundant information, resolving inconsistencies, and structuring the data in a way that facilitates easy access and interpretation for further investigation.

3.8.1 Standardisation and Normalisation of data

Standardization and normalization are techniques used in data pre-processing, which are employed to improve the performance and accuracy of machine learning models. Standardization involves scaling the features of a dataset. This process ensures that all variables contribute equally to the model, preventing any feature from dominating due to differences in their scales. On the other hand, the normalization focuses on transforming the values of features. By the process, it becomes less sensitive to the scale of the input, which helps with convergence during training and enhances the model's ability to handle types of inputs. Both standardization and normalization play roles in creating a ground for various features allowing machine learning algorithms to uncover meaningful patterns and relationships, within the data ultimately leading to more accurate and reliable predictions.

The data was standardized to exclude the animals that did not complete their lactation (i.e. earlier of either completing 305 days or being dried) on the farm. Further, the animals with extremely short or extremely long lactations as described by Lalhmingmawii (2021) for buffaloes and Mehta (2020) for crossbred cattle were also

excluded from the study. Lastly the milk yield data was also verified for its near normality by their descriptive statistics.

3.8.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step performed before pre-processing. It involves visual and statistical analysis of the dataset to gain insights into its distribution, central tendencies, and variability. EDA aids in identifying patterns, outliers, and potential relationships between variables, providing a foundation for informed pre-processing decisions.

The data Pre-processing steps involved deriving the daily milk yield data from morning and evening milk yield. The data were arranged into days of lactation from 1 to 305, segmenting the dataset for detailed analysis. The standard test days in monthly intervals (5, 35, 95, 125, 155, 185, 215, 245, 275, and 305) were specifically chosen for further investigation, potentially capturing critical points in the lactation curve.

3.8.3 Least Squares Analysis

The data analysis involved conducting a mixed model analysis using the Least-Square Maximum Likelihood program (Harvey, 1990) to investigate the influence of genetic and non-genetic factors on both 305-days yield and test day milk yield records of first lactation Murrah buffaloes. The models employed for this analysis were as follows:

$$Y_{ijkm} = \mu + S_i + P_j + L_k + e_{ijkm}$$

where

Y_{ijkm} = 305DMY of the mth individual of ith season of jth period and k^{th} parity

μ = Mean

S_i = Fixed effect of ith Season (i=1 to 4)

P_j = Fixed effect of jth Period (j=1 to 3)

L_k = Fixed effect of kth Parity group (k=1to 5)

e_{ijkm} = Random error

3.9 Evaluation of Machine Learning models for prediction of 305DMY

3.9.1 Training and testing of ML model using STDMY for predicting 305DMY

The process began by categorizing the total number of animals in each breed into three sets: training, testing, and validation, maintaining a proportion of 8:1:1, equivalent to 80%, 10%, and 10% of the total population, respectively. This stratification ensures a balanced and representative distribution for model development and evaluation Figure 3.1.
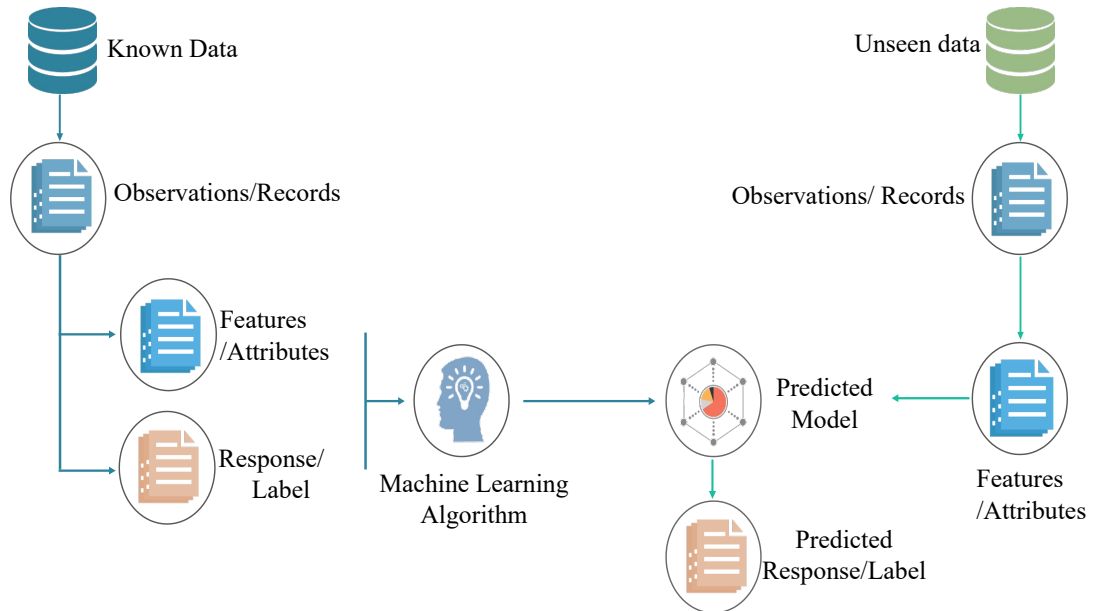

Figure 3.1: Working of Supervised Machine learning Model

3.9.2 Training and testing of ML model using RTDMY for predicting 305DMY

To facilitate the functioning of the supervised ML model, a novel method was employed to generate a subset of a thousand (1000) RTDMY sets from per lactation, implementing the following conditions.

● First random test day in each set falls between 5th and 15th day of lactation

● Subsequent random test days were randomly picked with constraint that the gap between them was no less than 15 days and no more than 45 days

● The milk yield corresponding to the random test day in the set was treated as random test day milk yield

These records originated from the daily milk yield of the animals specifically within the training set. The heart of the process involved training the ML model. This was achieved by employing ML algorithms and utilizing the RTDMY records derived from the training set of animals. The model was trained to predict 305DMY. This

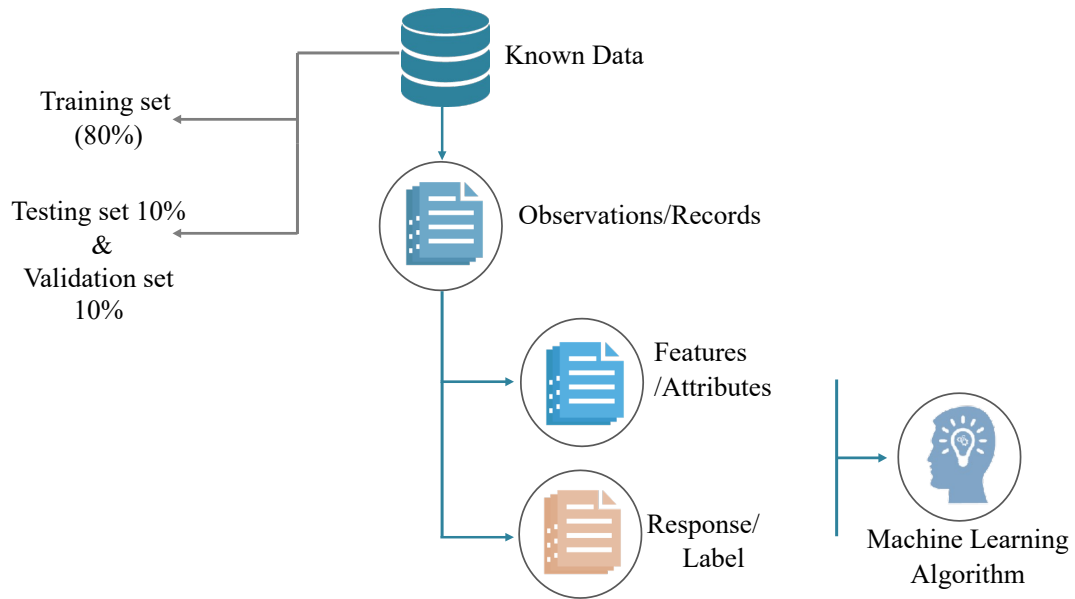comprehensive training phase allowed the model to grasp patterns and relationships within the data Figure 3.2.

Figure 3.2: Training, Testing, and Validation of Model

To further refine the model's predictive capabilities, various non-genetic factors were also considered as inputs for training the models. These factors included period, season and parity. By incorporating these additional variables, the aim was to improve the model's ability to predict the 305-days milk yield accurately. This optimization step enhanced the model's performance by considering multiple influencing factors in the prediction process.

3.9.3 Statistical analysis for ML models

The predicted 305DMY milk yields were compared to the observed 305DMY of the animals to determine the accuracy of the trained model. The training and analysis were conducted using R and Python environment. The statistical analyses were conducted in R version 3.3.1 (R Core Team 2016), employing specific packages for various methods.

The following r packages were used for training and testing of ML model.

The `caret` package in R by Kuhn (2008) a comprehensive tool designed for "Classification And REgression Training." It streamlines the complex processes involved in machine learning tasks, providing a unified interface for data pre-processing, model training, tuning, and evaluation.

The `randomForest` package in R by Liaw and Wiener (2002) was dedicated for constructing and analysing random forests, an ensemble learning method. This approach builds multiple decision trees during training and combines their results to improve predictive accuracy. The package's tunable parameters allow for optimal model performance in both classification and regression tasks.

The `XGBoost` package implements the eXtreme Gradient Boosting algorithm, known for its efficiency and predictive prowess. With features such as regularization techniques, tree pruning, and parallel computation, `xgboost` is a robust tool. It was provided with the flexibility to fine-tune hyperparameters, enabled to achieve optimal model performance.

R's `neuralnet` package instrumental in creating and training artificial neural networks (ANN). These versatile ML models, inspired by the human brain, support the construction of multi-layer perceptrons for both classification and regression tasks.

The fundamental `lm` function in R, part of the base package, is widely used for fitting linear models. This statistical method models relationships between a dependent variable and independent variables, offering interpretable coefficients and making linear models.

The `e1071` package introduces the `svm` function, enabling the implementation of Support Vector Machines (SVM) for regression. SVM, renowned for its power in both classification and regression, seeks to find a hyperplane representing the data distribution. The package provides flexibility in selecting kernels and tuning parameters for optimal results

R's `kernlab` package includes the `ksvm` function, designed for Support Vector Machines in regression (SMOreg). Particularly used for kernel-based methods, this package offers options for tuning parameters and selecting various kernel functions, enhancing its adaptability to diverse datasets.

The `tidymodels` framework integrates multiple modeling packages, including `caret`, promoting a tidy and systematic approach to the modeling workflow and it facilitated seamless model building and evaluation.

Essential for data manipulation (`dplyr`) and visualization (`ggplot2`), these packages was used in the initial stages of data pre-processing. Their functions aid in

creating insightful visualizations, enhancing the understanding of dataset characteristics and model outputs.

3.10 Comparative efficiency of ML model with conventional LC model

Following the training and testing of ML models using STDMY and RTDMY, their efficacy was assessed and validated against two conventional LC methods - Gamma Incomplete function (Wood, 1967) and Polynomial Regression Function (Ali and Schaeffer, 1987). Initially, 11 STDMY were selected derived from the daily milk records at monthly intervals, starting from day 5 until day 305 of lactation of an individual animal. These records served as the basis for predicting the 305-days milk yield employing both Wood's model and Polynomial Regression Function (PRF). The lactation curve models, along with multiple regression, random regression, and mixed models using test day milk yield records, were applied to determine the 305-days milk yield.

- **Gamma type function (Wood, 1967)**

    $Yt = at^b e^{-ct}$

    Where, Y_t = average daily yield in the tth test day of lactation

    a = initial milk yield just after calving

    b = ascending slope parameter till the peak yield

    c = descending slope parameter

    t = length of time since calving

    The constants can be derived by solving the above equation after

    transformation on the log scale.

    $$InY_t = In(a) + bIn(t) - ct$$

- **Polynomial Regression function model (Ali & Schaeffer, 1987)**

    $Yt = a + bx + cx^2 + d \log(1/x) + e \log(1/x)^2$

    Where, Y_t = Average milk yield in t^{th} day of lactation

    a = Associated with peak yield

    b & c = Associated with the decreasing slope

    d & e = Associated with the increasing slopes

    X= t / 305

3.11. Evaluation matrices

The comparative evaluation of ML models against conventional LC methods for predicting the 305DMY was conducted. A goodness-of-fit test is a statistical method used to assess whether a given set of observed data aligns with the expected values predicted by a specific model (Salau and Jain 2021). Model fit statistics were compared across different models. All analyses were performed using SAS/R software. Both ML and LC are evaluated using the following "Goodness of fit" criteria.

3.11.1 Coefficient of determination

Coefficient of determination (R^2) represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one (Tatachar 2021)

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \overline{y})^2}$$

Where y_i– Observed values

$\hat{y}$– Predicted values

3.11.2 Adjusted R^2

Adjusted R^2 is a modified version of R^2, and it is adjusted for the number of independent variables in the model, and it will always be less than or equal to R^2. (Tatachar 2021)

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-p-1}\right]$$

Where R^2 = R squared

n = number of observations

p = number of the independent variables

3.11.3 Mean absolute error

Mean Absolute Error (MAE) calculates the mean absolute deviations between the observed and predicted value pairs across all samples. It's referred to as the L_1 loss function, which measures a non-directional error by averaging the magnitudes of differences.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}|$$

Where y_i– Observed values

$\hat{y}$– Predicted values

n = Number of observations

3.11.4 Mean Squared Error

Mean Squared Error (MSE) represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2$$

Where y_i = Observed values

$\hat{y}$ = Predicted values

n = Number of observations

3.11.5 Root Mean Squared Error

Root Mean Squared Error (RMSE) is the square root of Mean Squared error. It measures the standard deviation of residuals.

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2}$$

Where y_i = Observed values

$\hat{y}$ = Predicted values

n = Number of observations

3.11.6 Relative Absolute Error

Relative Absolute Error (RAE) is calculated by dividing the total absolute error by the absolute difference between the mean and the actual value (Kaliappan et al., 2021)

$$\text{RAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i - \overline{y}|}$$

$$\text{Where } \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Where y_i = Observed values

$\hat{y}$ = Predicted values

n = Number of observations

3.11.7 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) calculated by dividing the absolute difference between the actual and predicted values by the actual value. Lower MAPE values indicate better model performance. (Kaliappan et al., 2021)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%$$

Where y_i = Observed values

$\hat{y}$ = Predicted values

n = Number of observations

3.12 Estimation of genetic parameters for 305DMY

The REML estimates of the variance components were computed for non-genetic factors, and animals through mixed model analysis, separately for the observed and predicted 305DMY data from Murrah, and Nili Ravi buffaloes and crossbred cattle. These estimated variance components served as the basis for determining heritability and repeatability of the 305DMY.

3.12.1 Heritability

The heritability for the observed 305DMY and predicted 305DMY was estimated from the variance components estimates through the REML method in the BLUPF90 family of software (Misztal et al., 2004).

3.12.2 Repeatability

The repeatability for the observed 305DMY and predicted 305DMY was estimated from the variance components estimates through the REML method in the BLUPF90 family of software (Misztal et al., 2004).

3.13 Estimation of breeding value for 305DMY

The estimated breeding values (EBV) for all the animals were derived from the BLUP using the REML estimated variance-covariance matrix. To assess the consistency and validity of the breeding values, animal and sire rank correlation analyses were conducted, allowing for the comparison and association of breeding values obtained from observed 305DMY with those predicted 305DMY from both ML and conventional LC models.

3.13.1 Pearson's correlation

The EBV values obtained from BLUP for the observed and predicted 305DMY were used to calculate the Product Moment correlation aka Pearson's correlation. The

correlations were calculated separately for EBVs of animals and sires included in the data sets.

$$r = \frac{\sum(X-\bar{x})(Y-\bar{Y})}{\sqrt{(X-\bar{x})2}\sqrt{(Y-\bar{Y})2}}$$

Where

$\bar{x}$ = Mean of X variable

$\bar{Y}$ = Mean of Y variable

3.13.2 Spearman's rank correlation

The correlations between the rankings of the sires based on their EBVs was done using Spearman's rank correlation (Spearman, 1987):

$$r(s) = 1 - \frac{6\sum di2}{n(n2-1)}$$

Where,

$r(s)$ = Rank correlation coefficient

n = Number of sires under evaluation

d_i = Difference of rank between paired items under two methods.

3.14 Model Ranking

The models were ranked for their overall usability with the following criteria

- Predictive performance using STDMY: RMSE and MAE

- Predictive performance using RTDMY: RMSE and MAE

- Variance component preservation: Heritability and Repeatability

- Applicability for genetic selection: Correlation between EBV of animals and sires

By evaluating based on the above factors, the ranking aims to identify models that demonstrate superior predictive accuracy, and suitability for informing breeding decisions, thus guiding in selecting the most effective model for their dairy farming operations.

CHAPTER IV

RESULTS AND DISCUSSION

The present research aims to assess and validate both traditional lactation curve models and machine learning models for predicting 305-days milk yield in cattle and buffaloes. The approach involves the utilization of eleven (11) standard test day milk yield (STDMY) records viz. milk yield on the fifth day (TDMY5), thirty-fifth day (TDMY35), sixty-fifth (TDMY65), ninety-fifth day (TDMY95), hundred twenty-fifth (TDMY125), hundred fifty-fifth (TDMY155), hundred eighty-fifth (TDMY185), two hundred and fifteenth (TDMY215), two hundred and forty fifth (TDMY245), two hundred and seventy fifth (TDMY275), and three hundred and fifth (TDMY305) milk yield, as well as incorporating a vast set of random test days' milk yields (RTDMY) to estimate the 305-days milk yield (305DMY). The study involves a comprehensive comparison and evaluation of the efficacy between the traditional approach of using STDMY records and the novel method employing RTDMY records for predicting the 305-days milk yield. The milk yield records for both the standard and random days will be used in deploying the conventional lactation curve methods and various machine learning models for evaluating their relative predictive efficiency for predicting the 305-days milk yield.

The aim of the present study is to determine the approach yielding better accuracy in predicting the overall milk production over the 305-days period (305DMY). This comparative analysis further aims to provide insights into the potential advantages and limitations of each method, contributing to the refinement of predictive models for more accurate and reliable estimation of 305-days milk yield in the studied populations to employ this information for genetic evaluation, for potentially identifying genetically superior animals.

The dataset used for the study spans over the period of nine years i.e. from year 2014 to 2022, encompassing over 1300 lactations from crossbred cattle and two different breeds of buffaloes as Murrah and Nili Ravi. The conventional and machine learning approaches have been applied separately for each breed. The factors influencing the lactation curve, mixed model analysis has been employed using the 'R' packages.

Genetic parameters have been estimated using the 'Restricted Maximum Likelihood' (REML) method and the breeding values of sires were determined using 'Best Linear Unbiased Prediction' (BLUP) using the variance complement estimates from REML. Both the REML and BLUP were deployed using the BLUPF90 suite of software (Misztal et al., 2004). This comprehensive approach aids in the effective genetic evaluation of superior animals, providing valuable insights for the selection of high-quality breeding stock. In the context of the present study, the REML and BLUP were, however, used for the comparison of genetic parameters and breeding values being predicted from the predicted 305DMY using different conventional and machine learning approaches.

4.1 Data structure

The study involved the analysis of 1662 lactation records from 940 dairy animals maintained at the Directorate of Livestock Farm, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana during the period from 2014 to 2022. Table 4.1 provides a concise overview of the data used in the current research. The 305DMY of dairy Murrah, Nili Ravi buffaloes and crossbred cattle (CB cattle) was estimated by employing both two different traditional and six different machine learning approaches, utilizing the standard test day milk yield records (TDMY5, TDMY35, TDMY65, TDMY95, TDMY125, TDMY155, TDMY185, TDMY215, TDMY245, TDMY275, and TDMY305) and thereafter by random test day milk yield (RTDMY) records.

Table 4.1: Summary of recorded data for three different breeds

Data	Number			Total
	Murrah	**Nili Ravi**	**CB Cattle**	
Number of animals	330	287	323	940
Number of lactations	648	486	528	1662
Duration	2014-2022	2014-2022	2014-2022	(9 years)

The study included animals with varying distribution of lactation length, ranging from short to prolonged periods; the extremely short and long lactating animals were excluded as described by Lalhmingmawii (2021) and Mehta (2020). After processing, the data structure was modified in Table 4.2.

Table 4.2: **Summary of final data structure under study for three different breeds**

Data	Number			Total
	Murrah	**Nili Ravi**	**CB Cattle**	
Number of animals	304	199	264	767
Number of lactations	583	359	401	1343
Duration	2014-2022	2014-2022	2014-2022	9 (years)
Sire	89	24	52	165
Dam	139	57	130	326

4.2 Descriptive statistics of milk yield in Murrah, Nili Ravi, and Crossbred cattle

4.2.1 Statistical summary for the 305DMY

This section provides insights into the lactation performance of Murrah, Nili Ravi, and crossbred cattle under the study. The mean 305 days milk yield, while including the regular and the segregated herds of the Directorate of Livestock Farms; was 2091.51±29.58 kg for Murrah, 1651.99±38.35 kg for Nili Ravi, and 3272.68±82.57 kg for crossbred cattle. The coefficient of variation reflected a moderate to high variability in the 305DMY trait, indicating 33.60% for Murrah, 43.93% for Nili Ravi, and 50.460% for crossbred cattle (Table 4.3).

Table 4.3: Descriptive statistics for 305DMY in Murrah, Nili Ravi buffaloes and crossbred cattle

Parameters	Murrah	Nili Ravi	CB Cattle
Number of lactations	583	359	401
Mean	2088.25	1651.99	3272.68
Standard error	29.07	38.36	82.57
Median	2102.60	1583.55	3224.00
Standard deviation	701.85	725.78	1651.40
Skewness	0.16	0.58	0.42
Kurtosis	0.24	0.35	0.11
Coefficient of variation	33.61	43.93	50.46

Lalhmingmawii (2021) also observed a similar value of 2060.27±23.64 kg. Meanwhile, Malhotra (2014) documented a mean of 879.4±18.85 kg for the first 305-days milk yield on the same animal farm. Jakhar et al. (2016) reported an average 305-days milk yield of 2060.93±20.22 kg, and Jamuna et al. (2015) found an average of 2078.20±31.21 kg in Murrah buffaloes. Pawar et al. (2012) estimated a higher yield of 2147.6±87.06 kg, while Thiruvenkadan et al. (2010) reported a lower average milk yield of 1616.3±39.6 kg.

Lalhmingmawii (2021) reported that the unadjusted mean for total lactation milk yield was found to be 2229.604±28.785 kg, a figure comparable to the reports of Pawar et al. (2012) as 2229.8±93.7 kg for Murrah buffaloes. Malhotra (2014) reported an average first lactation milk yield of 2118.91±26.67 kg while Rana (2020) reported it as 2150.03±13.19 kg for the identical herd of Murrah buffaloes. The 305DMY for Murrah was reported to be lower by other researchers, such as 1723.71±77.64 kg by Rao and Kumari (2021), 1877.29±45.59 kg by Pendor (2015), 1865±10.83 kg by Narayan (2011), and 1686.2±44.4 kg by Thiruvenkadan et al. (2010).

In the present study, the observed mean 305DMY for crossbred cattle was determined to be 3272.682±82.56 kg, with a coefficient of variation of 50.46%. This recorded value closely aligns with previous findings from Singh (2006), Kokate et al. (2014), Divya et al. (2014), and Singh et al. (2015), which reported estimates of 3197 kg, 3,068 kg, 3110 kg, and 3123 kg, respectively, for Karan Fries cattle, which is a composite cross of Holstein Friesen with indigenous cattle breed Tharparkar.

However, the descriptive statistics presented in Table 4.3, cannot properly be compared to other studies, as it has not been standardized as per the norms of excluding the potentially abnormal lactations of the segregated herds and the regular herd. Therefore, descriptive statistics of the herd milk yield performance were further calculated and presented in Table 4.4 to offer a comprehensive overview of the distribution and variability in 305DMY within each breed, after standardization of the records and removal of segregated herds data. The standardization norms involved an exclusion of all lactation from the animals from segregated herds, all lactations falling short of 150 days of lactation or having less than 1500 kg of milk yield.

Table 4.4: Descriptive statistics for standardized 305DMY in Murrah, Nili Ravi buffaloes and crossbred cattle

Parameters	Murrah	Nili Ravi	CB Cattle
Number of lactations	335	199	335
Mean	2283.98	2166.91	3731.80
Standard error	33.31	37.55	76.40
Median	2281.80	2034.10	3634.40
Standard deviation	609.76	529.75	1398.41
Skewness	0.46	1.27	0.76
Kurtosis	1.00	2.25	0.86
Coefficient of variation	26.70	24.45	37.47

4.2.2 Statistical summary for STDMY

A detailed descriptive statistics for the STDMY records for Murrah buffaloes, Nili Ravi buffaloes, and crossbred cattle is presented in Tables 4.5, 4.6, and 4.7, respectively.

Table 4.5: Descriptive statistics for STDMY in Murrah buffaloes

Traits	No. of observations	Mean	Standard Deviation	Standard error	Coefficient of variation
TDMY5	540	7.35	2.67	0.11	36.34
TDMY35	540	10.51	3.24	0.14	30.79
TDMY65	540	10.36	3.13	0.13	30.23
TDMY95	540	9.62	2.63	0.11	27.34
TDMY125	539	8.75	2.26	0.10	25.89
TDMY155	537	7.98	2.31	0.10	28.96
TDMY185	507	7.13	2.45	0.11	34.31
TDMY215	457	6.45	2.39	0.11	37.03
TDMY245	408	5.41	2.50	0.12	46.22
TDMY275	303	4.74	2.55	0.15	53.72
TDMY305	116	3.79	1.93	0.18	51.00

These tables from 4.5, 4.6 and 4.7 provide a comprehensive overview of key parameters such as mean, standard deviation, standard error, and coefficient of variation; offering insights into the distribution and variability of test day yields across the specified period for each species. These descriptive statistics serve as a valuable

resource for understanding the production patterns and characteristics of lactating Murrah buffaloes, Nili Ravi buffaloes, and crossbred cattle within the scope of the study.

From Table 4.5, it was evident that TDMY35 (35th day of lactation) exhibited the highest mean value of 10.50 for monthly test day milk yield, while the lowest was observed in TDMY305 (305th day of lactation) in Murrah buffaloes. Overall, there was an initial increase in standard test day milk yield until TD-3, followed by a gradual decline towards the conclusion of lactation. The coefficient of variation for standard test day milk yields varied between 30.22% and 53.72% in Murrah buffaloes. Further, there is a decline in the number of lactations with later test days after TDMY185 and a steep decline after TDMY 275. Singh et al. (2015) reported that the average monthly test day milk yields (MTDMY) showed an upward trend, starting at 5.91±0.13 kg on TD1 (15th day of lactation), reaching its peak at 7.41±0.12 kg on TD3 (75th day of lactation), and then gradually decreasing to 3.11±0.15 kg on TD10 (285th day of lactation) in Murrah buffaloes. This pattern is consistent with findings by Geetha et al. (2006), Chakraborty et al. (2010), and Patil et al. (2012), who reported peak MTDMY values of 7.15±0.15 kg, 8.11±0.25 kg, and 8.02±0.09 kg, respectively, occurring on TD3. The trends in STDMY from the present study infer an earlier and higher peak yield, compared to the above-mentioned reports from the earlier authors.

Table 4.6: Descriptive statistics for STDMY in Nili Ravi buffaloes

Traits	No. of observations	Mean	Standard Deviation	Standard error	Coefficient of variation
TDMY5	301	6.65	2.54	0.13	38.23
TDMY35	301	8.92	2.75	0.15	30.82
TDMY65	301	8.77	3.03	0.16	34.54
TDMY95	301	7.92	2.91	0.17	36.75
TDMY125	299	7.20	3.04	0.17	42.18
TDMY155	293	6.56	2.92	0.17	44.50
TDMY185	258	5.86	2.73	0.17	46.59
TDMY215	222	4.92	2.70	0.18	54.79
TDMY245	160	4.19	2.97	0.22	70.85
TDMY275	104	4.24	2.47	0.24	58.26
TDMY305	67	3.60	2.51	0.29	69.91

The highest mean for STDMY among Nili Ravi buffaloes occurred in TDMY35 as 8.92±0.15 kg, while the lowest milk was recorded in TDMY305 as 3.60±0.29 kg. The monthly test day milk yield exhibited an increase until TD-2 (35th day of lactation), followed by a gradual decline towards the conclusion of lactation for Nili Ravi buffaloes. The coefficient of variation for weekly test day milk yields varied between 30.82% and 70.85% (Table 4.6).

In Table 4.7, presenting descriptive statistics for STDMY in crossbred cattle, it was noted that the highest milk was observed in TD-2 as 16.23±0.36 kg followed by a gradual decline towards the later lactation. The number of animals lactating also declined gradually beyond the sixth test day and showed a sharp decline after the ninth and tenth test days. The coefficient of variation in cattle ranged from 44.66% to 67.77%.

Table 4.7: Descriptive statistics for STDMY in crossbred cattle

Traits	No. of observations	Mean	Standard Deviation	Standard error	Coefficient of variation
TDMY5	401	13.40	6.49	0.32	48.44
TDMY35	401	16.23	7.25	0.36	44.67
TDMY65	401	15.42	7.47	0.37	48.42
TDMY95	401	14.47	6.93	0.35	47.88
TDMY125	374	13.35	6.60	0.34	49.45
TDMY155	344	12.23	5.82	0.31	47.63
TDMY185	316	11.43	5.48	0.31	47.93
TDMY215	287	10.54	5.35	0.32	50.74
TDMY245	248	9.30	4.95	0.31	53.18
TDMY275	201	8.07	4.62	0.33	57.22
TDMY305	153	6.89	4.67	0.38	67.77

4.3 Non-genetic factors affecting the dairy performance traits

The effects of season and period of calving and parity were evaluated on 305DMY in Murrah, Nili Ravi and crossbred cattle. It was found that all three non-genetic factors significantly (p<0.05) affected the 305DMY in Murrah buffaloes. In Nili Ravi buffaloes, however, only effect of season was found significant (p<0.05). Whereas

the effects of season and period of calving were found significant (p<0.05) for 305DMY of crossbred cattle.

4.3.1 Effect of season of calving on 305DMY

The effect of the season was found to be significant for all three herds under study (p<0.05). The least squares mean values and their corresponding standard errors for the impact of seasons on 305DMY are detailed in Table 4.8. The observed least squares mean for 305DMY reached its peak at 2326.35±59.09 kg during the winter season, which was significantly (p<0.05) higher than rainy (2086.36±59.03 kg), whereas 305DMy for Spring (2281.87±60.27kg), Autumn (2193.40±59.05kg), Summer (2133.52±66.75 kg) differed non-significantly in Murrah buffaloes.

Notably, the results show variations in milk yield across different seasons in Nili Ravi buffaloes. Specifically, the milk yield is significantly higher (p<0.05) during Winter (2038.62±92.39 kg), during Rainy (1721.74±75.13 kg); whereas 305DMY in Spring (1944.08±69.30 kg), Summer (1905.03±94.42 kg), and Autumn (1761.70±84.38 kg), did not differ significantly from any of the seasons.

Table 4.8: Effect of Season of calving on 305DMY

Season	LSM ± SE		
	Murrah	**Nili Ravi**	**CB Cattle**
Autumn (16th Sept to 30th Nov)	2193.40[ab]±59.05	1761.70[ab]±84.38	3457.85[ab]±164.22
Rainy (1st Jul to 15th Sept)	2086.36[b]±59.03	1721.74[b]±75.13	3225.40[b]±192.78
Spring (1st Feb to 15th Apr)	2281.87[ab]±60.27	1944.08[ab]±69.30	3536.93[ab]±200.22
Summer (16th Apr to 30th Jun)	2133.52[ab]±66.75	1905.03[ab]±94.42	2841.04[b]±266.76
Winter (1st Dec to 31st Jan)	2326.35[a]±59.09	2038.62[a]±92.39	3974.53[a]±179.15

The values having different superscripts within a breed differ significantly (p<0.05)

The least squares means for 305DMY was observed to be significantly higher for winter (3974.53±179.15 kg), than summer (2841.04±266.76) and rainy (3225.40±192.78) season in crossbred cattle. For first lactation 305DMY, Singh et al.

(2015) reported the highest value (3270 kg) during the autumn season, followed by rainy (3139 kg), winter (3138 kg), and summer (3095 kg), and Rashia (2010) who reported the highest first lactation 305DMY in autumn (3059 kg) and the lowest during the summer season (2901 kg) in Karan Swiss cattle. In contrast, Divya et al. (2014) observed a non-significant effect of the calving season on 305DMY in Karan Fries cattle.

4.3.2 Effect of period of calving on 305DMY

Table 4.9 summarizes the effect of the period with the least squares means and their corresponding standard errors (SE) for 305DMY across three herds -Murrah and Nili Ravi buffaloes, and crossbred cattle - over three consecutive periods: 2014-2016, 2017-2019, and 2020-2022. For Murrah, the 305DMY was highest at 2273.80±58.01 kg in 2014-2016. Nili Ravi exhibited no significant differences in 305DMY across the periods. In the case of crossbred cattle, there was a significant (p<0.05) decrease from 3818.59±237.76 kg in 2014-2016 to 2916.00±147.27 kg in 2017-2019, followed by an increase to 3486.86±136.64 kg in the period 2020-2022. These trends suggest notable fluctuations in 305DMY across seasons and periods for crossbred cattle.

Table 4.9: Effect of Period on 305DMY

Period	LSM ± SE		
	Murrah	**Nili Ravi**	**CB Cattle**
2014-2016	2273.80[a]±58.01	1795.66±68.30	3818.59[a] ±237.76
2017-2019	2238.43[ab]±44.52	1859.96±62.14	2916.00[b] ±147.27
2020-2022	2100.66[b]±44.10	1967.07±63.46	3486.86[a] ±136.64

The values having different superscripts within a breed differ significantly (p<0.05)

Dalal et al. (1999), Jadhav et al. (1998), and Singh et al. (1995) reported a highly significant (P<0.01) impact on the period of calving 305DMY in various crossbreds. Similar findings were also reported by Taneja and Sikka (1981) in Tharparkar, Sahiwal, and Red Sindhi, as well as by Rao et al. (1984) in Ongole, Tharparkar, Malvi, and crossbreds.

4.3.3 Effect of Parity on 305DMY:

Table 4.10 presents the least squares means (LSM) along with their standard errors for the non-genetic factor, Parity, influencing 305DMY in Murrah buffaloes. The LSM values for each parity level are reported with their respective standard errors. Notably, the results show a significant increase in LSM from first parity to second parity, which was no different to third and fourth parity, Subsequently, there is a decrease in 305DMY for the group fifth parity and more in Murrah buffaloes. Whereas, no significant differences among parity groups were observed in Nili Ravi and crossbred cattle (Table 4.10).

Table 4.10: Effect of Parity on 305DMY

Parity	LSM ± SE		
	Murrah	**Nili Ravi**	**CB Cattle**
First	2054.89[c]±43.67	1830.82±65.24	3690.05±145.78
Second	2314.06[ab]±53.31	1995.82±70.78	3589.55±148.13
Third and fourth	2346.79[a]±52.08	1957.49±66.35	3361.70±157.37
Fifth and more	2101.46[bc]±76.00	1712.78±100.36	2987.30±352.03

The values having different superscripts within a breed differ significantly (p<0.05)

Kumar et al. (2021) said that there was no significant impact observed for non-genetic factors such as season, period of lactation, and parity on the 305 DMY. Additionally, third-party animals displayed the highest 305 DMY, whereas cows in the fifth or higher parity had the lowest 305 DMY. This aligns with findings by Verma et al. (2016), who also observed no significant impact of parity on 305 days, although they noted that the 3rd parity exhibited the highest yield. In contrast, Bajwa et al. (2004) reported differing results, suggesting that milk yield increases with an increase in parity. According to Bajwa et al.(2004), the highest production was achieved around the 4th and 5th parities, followed by a declining trend thereafter.

Parity demonstrated a noteworthy effect (p<0.05) on the monthly milk yield by Kumar et al. (2021) in Sahiwal cattle. Also, similar findings were reported by Jingar et al. (2014) and Japheth et al. (2015), who observed a significant influence of parity (P<0.01) on monthly milk yield in Karan Fries cows. Additionally, Kumari (2019) and Sinha (2019) found a significant impact of parity on milk yield in Sahiwal cows. In

Sahiwal cattle, the monthly milk yield exhibited an upward trend up to the 4th parity, followed by a decline with increasing parity, as noted in the studies mentioned.

4.4 Machine learning models for predicting 305DMY using STDMY

A total of six Machine Learning (ML) models were evaluated to predict 305DMY using STDMY which included K Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Random Forest (RF), Sequential Minimal Optimization Regression (SMOreg), Machine learning Linear Regression Model (MLR), and eXtreme Gradient Boosting (XGB). The assessment compared the performance of these models based on goodness of fit criteria for the training and validation sets.

The training and testing of the dataset were conducted in the R environment simultaneously for all the ML models. A meticulous division of data as 80% during training and 10% during testing was done. Additionally, validation was performed using 10% of the data for each breed, ensuring robustness and reliability in the model. This concurrent method, in R, enabled us to evaluate the models' performance ensuring its ability to generalize well to data sets. By allocating 10% of the dataset for validating each breed the model's ability to classify across different categories was thoroughly assessed, enhancing its overall reliability and applicability.

4.4.1 Training and Predicting 305DMY using STDMY for Murrah buffaloes

The results showed (Table 4.11) that the Machine learning Linear Regression Model (MLR) achieved the highest R (correlation) value of 0.9863 followed closely by ANN (0.9830) and XGB (0.9738) indicating strong correlation of predicted values with actual 305DMY. MLR also performed well in R-squared and Adjusted R-squared values demonstrating efficiency in capturing variance while considering predictors, closely followed by ANN and XGB. Additionally, MLR had the Mean Absolute Error (MAE) at 79.1808 showing accuracy in predicting absolute differences, and excelled in Mean Absolute Percentage Error (MAPE) with a minimal value of 4.01598% reflecting precision, in predicting percentage errors.

Moreover, MLR demonstrated the performance, with the minimum Root Mean Square Error (RMSE) value of 114.7233 highlighting its effectiveness in reducing prediction inaccuracies followed closely by XGB and ANN. Comparative measures like Relative Absolute Error (RAE) and Relative Root Square Error (RRSE) further

supported the results. Although other models also showed good performances; a thorough evaluation consistently ranked the MLR model, followed by XGB and ANN and RF as the performer based on various assessment criteria showcasing its reliability, in both explanatory and predictive functions.

Table 4.11: Goodness of fit of ML algorithms for predicting 305DMY using STDMY in Murrah buffaloes

Parameters	KNN	ANN	RF	SMOreg	MLR	XGB
R	0.9709	0.9830	0.9709	0.9596	0.9863	0.9738
R^2	0.9427	0.9662	0.9427	0.9208	0.9728	0.9483
Adjusted R^2	0.9422	0.9659	0.9422	0.9201	0.9726	0.9479
MAE	1851.8578	135.1378	118.8621	112.1900	79.1808	118.7779
MAPE	90.0655	8.0327	6.1246	5.4842	4.01598	5.9283
RMSE	1937.3052	172.0332	175.7308	202.0101	114.7233	159.3133
RAE	3.4324	0.2504	0.2203	0.20794	0.1467	0.2201
RRSE	2.7998	0.2486	0.2540	0.2919	0.1658	0.2302

4.4.2 Training and Predicting 305DMY using STDMY for Nili Ravi buffaloes

The evaluation of six ML models for predicting 305DMY based on STDMY in Nili Ravi buffaloes (Table 4.12) revealed that MLR, RF, and XGB emerged as the top three performers. MLR demonstrated the highest R, R^2, and Adjusted R^2 values, showcasing superior explanatory power and accuracy. RF exhibited competitive results across various metrics, benefitting from its ensemble nature and ability to capture complex relationships. XGB showcased strong performance, particularly in terms of R and R^2, leveraging boosting for iterative improvement.

The MLR stands out for its overall excellence in explanatory and predictive capabilities, RF for its robustness, and XGB for its adaptability and boosting mechanism. Each model presents a valuable option for predicting 305DMY in Nili Ravi buffaloes.

In contrast, K-Nearest Neighbors (KNN) exhibited comparatively lower performance in predicting 305DMY in Nili Ravi buffaloes using STDMY. The KNN showed lower but comparable R, R^2, and Adjusted R^2 values, indicating slightly reduced explanatory power compared to the top-performing models. However, KNN

demonstrated very high Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) values, suggesting less precise predictions and higher overall errors. These lower-performing models underscore the importance of selecting appropriate algorithms, with MLR, XGB, and RF proving to be more effective choices for accurate and reliable predictions in the context of Nili Ravi's buffalo milk yield prediction.

Table 4.12: **Goodness of fit of ML algorithms for predicting of 305DMY using STDMY in Nili Ravi buffaloes**

Parameters	KNN	ANN	RF	SMOreg	MLR	XGB
R	0.9754	0.9638	0.9772	0.97281	0.9875	0.9789
R^2	0.9514	0.9289	0.9550	0.94637	0.9751	0.9583
Adjusted R^2	0.9506	0.9279	0.9543	0.9455	0.9748	0.9577
MAE	1521.5132	147.3982	106.9202	103.8218	75.1338	99.9082
MAPE	91.9079	9.0774	5.9593	6.2612	4.6891	5.8126
RMSE	1642.6530	211.8832	160.0852	176.2554	111.8572	142.3964
RAE	2.7158	0.2630	0.1908	0.1853	0.1341	0.1783
RRSE	2.3602	0.3044	0.2300	0.2532	0.1607	0.2045

4.4.3 Training and Predicting 305DMY using STDMY for crossbred cattle

The model efficiency parameters for assessing ML models to predict 305DMY in crossbred cattle using STDMY measures, are shown in Table 4.13. MLR Model shines with the R-value of 0.9871, R^2 value of 0.9745, and Adjusted R^2 of 0.9742 values indicating its explanatory power and precision. Artificial Neural Networks (ANN) follow closely with performance metrics, like R (0.9833), R^2 (0.9668), and Adjusted R^2 (0.9664) showcasing their ability to capture data relationships effectively. XGBoost (XGB) demonstrates performance in terms of R (0.9758), R^2 (0.9522), and adjusted R^2 (0.9516) positioning it as an option for forecasts. Random Forest (RF) also delivered comparable results with values for R (0.9748), R^2(0.9503), and Adjusted R^2(0.9497) offering reliable predictions.

The least RMSE was observed for MLR followed by XGB and RF, while the least MAE was for MLR, followed by SMOreg, RF, and XGB. SMOreg however has a high RMSE, making it less reliable for the prediction of extreme values further

discouraged by comparatively lower R^2. On the other side, K Nearest Neighbor (KNN) showed weaker performances with higher error rates observed in their metrics.

In conclusion, MLR followed by RF and XGB emerge as leading models, closely followed by XGB and RF delivering accuracy, in predicting 305DMY in dairy crossbred cattle; while KNN and SMOreg exhibit the least efficacy within this framework.

Table 4.13: **Goodness of fit of ML algorithms for predicting of 305DMY using STDMY in crossbred cattle**

Parameters	KNN	ANN	RF	SMOreg	MLR	XGB
R	0.9636	0.9833	0.9748	0.9442	0.9871	0.9758
R^2	0.9285	0.9668	0.9503	0.8915	0.9745	0.9522
Adjusted R^2	0.9276	0.9664	0.9497	0.8901	0.9742	0.9516
MAE	3122.4017	298.4983	246.1364	233.4109	177.1694	258.9736
MAPE	95.4852	14.0803	8.2078	7.1673	6.0785	8.4469
RMSE	3466.5978	387.1738	374.6540	532.2520	256.2250	353.4422
RAE	2.4210	0.2314	0.1908	0.1809	0.1373	0.2008
RRSE	2.1791	0.2433	0.2355	0.3345	0.1610	0.2221

The scatter plots with histograms provide a comprehensive visual analysis of the predicted versus observed values of 305DMY in Murrah buffaloes (Figure 4.1), Nili Ravi buffaloes (Figure 4.2), and crossbred cattle (Figure 4.3) utilizing STDMY for six ML models.

The dot-plot provides a visual representation of the observed and predicted yield across six ML models for Murrah buffaloes (Figure 4.4), Nili Ravi buffaloes (Figure 4.5), and crossbred cattle (Figure 4.6).

Notably, the dot-plots reveal a significant performance gap, particularly for the K-Nearest Neighbors (KNN) model in the prediction of 305DMY. In contrast to the other models, the dots representing KNN's predicted values widely deviated noticeably from the observed values, indicating a lack of accuracy in its predictions. This inconsistency implies that the KNN model struggles to capture the inherent patterns and relationships within the data for all three breeds.

On the flip side, the dots representing predictions from the other models (ANN, RF, SMOreg, MLR, XGB) are closely clustered, aligning well with the observed values. This coherence suggests that these models, whether it be the machine learning Linear Regression model, Random Forest, or others, demonstrate a higher level of accuracy in predicting 305DMY for Murrah buffaloes, Nili Ravi buffaloes, and crossbred cattle.

The dot-plots underscores the importance of thoughtful model selection, with a clear indication that KNN may not be the most suitable choice for precise 305DMY predictions.

4.5 Training Machine learning model for predicting 305DMY with random test day milk yield (RTDMY) records

To address the rigidness in STDMY records method in predicting the 305DMY, an innovative approach of random test days milk yield (RTDMY) records was utilized for predicting the 305DMY. For this purpose, a subset of a thousand combinations of random test days was created, and milk yields corresponding to those 1000 random test days were drawn from each lactation for each breed, implementing specific conditions to ensure diversity, robustness, and reliability in the generated dataset. The conditions in drawing the random test days were set such that the first random test day in each set fell within the range of 5 to 15 days of lactation to not miss the very initial phase of lactation while avoiding the colostrum period. Another crucial aspect of the method was to enforce a consistent gap between consecutive test days within each set. Specifically, a constraint was applied to ensure that the gap remained no more than 40 days and, importantly, no less than 15 days. This condition aimed to emulate realistic scenarios in testing schedules, contributing to the authenticity of the generated dataset. This careful restriction was introduced to maintain variability in the dataset and avoid biases that might arise from too close or too farther apart random test days.

By incorporating these conditions, the methodology not only facilitated the creation of diverse and representative subsets but also introduced a level of flexibility along with temporal coherence, aligning with the practical constraints often encountered in the context of standard test day milk yield for different breeds. Subsequently, the milk yield records corresponding to the 1000 sets of random test days were drawn from each lactation under study.

During the analysis of 1000 sets of random test day milk yield (RTDMY) from each lactation to predict 305DMY using the earlier described six ML models, the outcomes of the training and predictions were obtained and retained along with the model efficiency parameters. The dot-plots in Figure 4.7 display the adjusted R^2 values, for three herds; Murrah, Nili Ravi, and crossbred cattle, respectively.

The data in the graph presents adjusted R^2 values for six machine learning models giving insights into how well they predict 305DMY in three specific three breeds; Murrah and Nili Ravi, crossbred cattle.

K Nearest Neighbor (KNN) consistently performs across breeds with adjusted R^2 values ranging from 0.9408 to 0.9457 for Murrah (Figure 4.8a), 0.9442 to 0.9533 for Nili Ravi (Figure 4.8b)and 0.9233 to 0.9286 for crossbred cattle (Figure 4.8c).

Artificial Neural Network (ANN) show reliable predictive accuracy with adjusted R^2 values from 0.9037 to 0.9766 for Murrah (Figure 4.9 a), 0.8595 to 0.9661 for Nili Ravi (Figure 4.9b) and 0.9053 to 0.9661 for crossbred cattle (Figure 4.9c). The range of adjusted R^2 values indicate how well the ANN model fits in predicting 305DMY variations for each breed.

Random Forest (RF) stands out with its abilities showing adjusted R^2 values ranging from 0.9236 to 0.9796 for Murrah (Figure 4.10a) and 0.92 to 0.98 for Nili Ravi (Figure 4.10b) and 0.90 to 0.96 for crossbred cattle (Figure 4.10c).

Sequential Minimal Optimization (SMO) regression produces a range of adjusted R square values, from 0.8829 to 0.9588 for Murrah (Figure 4.11a), 0.8473 to 0.97 for Nili Ravi (Figure 4.11b), and 0.8261 to 0.91 for crossbred cattle (Figure 4.11c).

The MLR Model also demonstrated good performance with adjusted R^2 values ranging from 0.94 to 0.98 for Murrah (Figure 4.12a), 0.8711 to 0.9840 for Nili Ravi (Figure 4.12b), and 0.92 to 0.9727 for crossbred cattle (Figure 4.12c).

Lastly, XGBoost (XGB) shows values ranging from 0.9337 to 0.9770 for Murrah (Figure 4.13a), 0.9302 to 0.9798 for Nili Ravi (Figure 4.13b), and 0.8949 to 0.9798 for Crossbred Cattle (Figure 4.13c).

It is evident that the K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Sequential Minimal Optimization regression (SMO), Random Forest (RF), eXtreme Gradient Boosting (XGB), and Linear Model (LM) are effective models for

making predictions of 305-days milk yield (305DMY) with respect to adjusted R^2 values. The consistency in obtaining adjusted R-square values across different breeds underline the adaptability and efficiency of these models in capturing the prediction of 305DMY. This consistency further supports their potential usefulness in field applications. These findings provide valuable insights into the performance and relevance of machine learning models across diverse breeds, particularly in the context of random test day milk yield. Further, it is evident that the RTDMY are performing as efficiently as the STDMY when used with different ML models. Moreover, the striking efficiency of RTDMY compared to STDMY across a spectrum of machine learning models serves as a robust validation of their performance consistency, emphasizing their reliability in diverse predictive scenarios. This observation reinforces the notion that RTDMY can serve as a dependable alternative to STDMY, offering comparable predictive accuracy across various breeds and machine learning methodologies. This consistency not only bolsters confidence in the reliability of RTDMY but also underscores the versatility and generalizability of the machine learning models employed, solidifying their applicability in real-world settings.

4.6 Conventional lactation curve method of predicting 305DMY using STDMY

Following the training of models using both STDMY and RTDMY records, their efficacy is rigorously assessed and validated against two conventional methods - Gamma Incomplete Function (Wood, 1967) and Polynomial Regression Function (Ali and Schaeffer, 1987).

Initially, to implement the two conventional lactation curve model, 11 standard test day milk yield (STDMY) records were extracted from the 5 to 305 days milk yield data similar to the machine learning model. These records were then employed to predict the 305DMY utilizing both the Gamma Incomplete Function aka (Wood model) and Polynomial Regression Function (PRF).

4.6.1 Gamma Incomplete function:

The evaluation of the Lactation Curve model particularly when using the Gamma incomplete function (Wood, 1967) approach to predict the 305DMY based on STDMY in three breeds (Murrah, Nili Ravi and crossbred cattle) revealed valuable insights (Table 4.14).

The Wood's model showed a fit across all breeds with high R values (Murrah: 0.9782, Nili Ravi: 0.9819 crossbred cattle: 0.9818), R^2 values (Murrah: 0.9570; Nili Ravi: 0.9643; crossbred cattle: 0.9639) and Adjusted R^2 values (Murrah: 0.9569; Nili Ravi: 0.9642; crossbred cattle: 0.9638) indicating its effectiveness, in prediction of 305DMY.

The mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) metrics further confirmed the accuracy of the model with values observed across all three breeds. Additionally, relative metrics like error (RAE) and relative root square error (RRSE) highlighted the precision of the model with lower values indicating a good fit (Table 4.21).

Table 4.14: Goodness of fit of Wood's lactation curve model for prediction of 305DMY using STDMY in three different breeds

Parameters	Murrah	Nili Ravi	CB Cattle
R	0.9782	0.9819	0.9818
R^2	0.9570	0.9643	0.9639
Adjusted R^2	0.9569	0.9642	0.9638
MAE	134.4258	116.6779	248.1348
MAPE	6.2847	6.8766	6.9888
RMSE	189.4404	173.6207	372.8199
RAE	0.2386	0.1994	0.1933
RRSE	0.2683	0.2391	0.2333

Kong et al. (2018) reported that the Wood's models have shown notable promise, closely mirroring the actual curves. Their effectiveness renders them valuable instruments for forecasting the 305-days yield in Chinese Holstein cattle, providing crucial insights for both managerial decisions and genetic evaluation. In their study, the Wood's model demonstrated a close approximation to the actual lactation curve across

all groups when utilizing 3 or 6 test day records. Notably, in the first lactation, the predicted curve closely aligned with the actual curve up to 175 days in milk (DIM) with 3 or 6 test day records, but subsequently exhibited a slight deviation, presenting lower values.

Similar observations have been reported in prior studies by Strabel et al. (2004), Macciotta et al. (2005), and Cankaya et al. (2011). Despite the Wood's model providing a physiologically grounded representation of the lactation curve, yielding a satisfactory fit as indicated by Cunha et al. (2010), it has been noted to slightly underestimate daily milk yield in various studies (Cilek & Keskin, 2008; Banu et al., 2012; Dohare et al., 2014). Despite this, the Wood's model serves as a valuable guide for further modelling efforts and contributes novel insights.

The scatter plots with histograms offer a comprehensive visual analysis of the predicted versus observed values of 305DMY in Murrah buffaloes (Figure 4.14), Nili Ravi buffaloes (Figure 4.15), and crossbred cattle (Figure 4.16) utilizing STDMY for the Wood's model using STDMY records.

The grouped scatter plot with marginal density plots offers a comprehensive visualization of the relationship between predicted and observed 305DMY values across different breeds Figure 4.17, while also providing insights into the distribution of values within each breed, all within the context of the Wood's Lactation curve model.

Figure 4.18 highlights the connection between the projected values using the Gamma's incomplete model for three breeds. The strong correlation coefficients noted for Murrah (0.978), Nili Ravi (0.982) and crossbred cattle (0.982) indicate relationships between the observed and predicted values for each breed. These significant correlation coefficients demonstrate the effectiveness of the Gamma's incomplete model in capturing and reproducing patterns within the data. The high correlation values for all three breeds confirm the reliability of the model in forecasting results providing insights into its performance and applicability, across animal populations.

4.6.2 Polynomial Regression Function

The Polynomial Regression function (PRF) method (Ali and Schaeffer, 1987) showed moderate to weak model fit statistics with R values (0.6595 for Murrah, 0.3176 for Nili Ravi and 0.6539 for crossbred cattle) indicating its struggle to capture data patterns effectively, when used for individual animals. The PRF as the Lactation Curve

model did not perform well for predicting the 305DMY for Murrah, Nili Ravi and crossbred cattle (Table 4.15).

The R^2 and Adjusted R^2 values indicated limited explanation of variance in 305DMY predictions highlighting the limitations of this method. Mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) values pointed to prediction errors showcasing the models lackluster precision. Metrics like error (RAE) and relative root square error (RRSE) further emphasized the method's shortcomings, especially in Nili Ravi and crossbred cattle. Ultimately these findings underscore the need for modelling approaches when predicting 305DMY in these specific dairy animal populations by using STDMY.

Table 4.15: Goodness of fit of Polynomial Regression Function method for prediction of 305DMY using STDMY in three different breeds

Parameters	Murrah	Nili Ravi	CB Cattle
R	0.6595	0.3176	0.6539
R^2	0.4349	0.1009	0.4276
Adjusted R^2	0.4339	0.0977	0.4261
MAE	223.6646	386.3577	597.9387
MAPE	12.8891	27.4159	25.6292
RMSE	581.2491	1204.0767	1561.6323
RAE	0.4360	0.7480	0.4553
RRSE	0.8986	1.8371	0.9569

Scatter plots, accompanied by marginal histograms, depicted the relationship between observed and predicted values of 305DMY in Murrah buffaloes reveal that the Polynomial Regression function exhibits poor predictive power (Figure 4.19).

The scatteredness observed in the plot further underscores the inefficiency of the model in accurately predicting 305DMY. The lack of a clear and consistent pattern in the scatter points suggests that the model struggles to capture the underlying dynamics of the milk yield data, resulting in suboptimal predictions. This observation emphasizes the limitations of the PRF in the context of 305DMY prediction for Murrah buffalo, urging caution and consideration of alternative modelling approaches for more accurate results.

Similar observations were noted in Nili Ravi buffaloes (Figure 4.20) and crossbred cattle (Figure 4.21), where scatter plots with marginal histograms of predicted versus observed values for 305DMY using the Polynomial Regression function revealed poor prediction power. The scattered distribution of points in both breeds indicates a lack of precision and efficiency in the model's ability to accurately predict 305DMY. This consistent pattern of inefficiency across multiple breeds underscores the limitations of the Polynomial Regression function, suggesting the need for alternative modelling approaches or adjustments to enhance predictive performance in Nili Ravi and crossbred cattle populations.

The Figure 4.22 presents a grouped scatter plot with marginal density plots, illustrating the connection between predicted and observed 305DMY values across various breeds. This visualization also offers insights into the distribution of these values within each breed, all under the framework of the polynomial regression curve model.

Figure 4.23 underscores the correlation between observed and predicted values using the Polynomial Regression method for three herds. The reported correlation coefficients for Murrah (0.889), Nili Ravi (0.318), and crossbred cattle (0.590) suggest varying degrees of association between the observed and predicted values for each breed.

The positive correlation for Murrah indicates a good relationship, affirming the Polynomial Regression model's effectiveness in capturing patterns specific to this breed. However, the lower positive correlation for Nili Ravi (0.318) implies a comparatively weaker association, while the moderate positive correlation for crossbred cattle (0.590) signifies a moderate level of relationship between observed and predicted values.

4.7.1 Gamma Incomplete Function method using RTDMY

In a similar vein, both the conventional Lactation Curve (LC) model underwent testing using 1000 sets of random test day milk yield (RTDMY) records per animal per lactation, showcasing highly variable and comparatively lower adjusted R^2 values when juxtaposed with the machine learning models derived using RTDMY's. Specifically, for the gamma incomplete model, the adjusted R^2 values hovered between the values of 0.54 to 0.69 in Murrah buffaloes (Figure 4.24a), 0.43 to 0.67 in Nili Ravi buffaloes (Figure 4.24b), 0.64 to 0.73 in crossbred cattle (Figure 4.24c).

4.7.2 Polynomial Regression Function using RTDMY

The adjusted R^2 values, in the polynomial regression analysis showed wide ranges for Murrah buffaloes Nili Ravi buffaloes, and crossbred cattle. Specifically, they ranged from adjusted R^2 values of 0.04 to 0.94 for Murrah buffaloes (Figure 4.25a), 0.17 to 0.96 for Nili Ravi buffaloes (Figure 4.25b), and 0.18 to 0.96 for crossbred cattle (Figure 4.25c).

This variability suggests that the predictive accuracy of the two conventional LC models are inconsistent and relatively poor when compared to advanced machine learning models. The fluctuations indicate that the Wood's and PRF model may encounter challenges in predicting outcomes across datasets and scenarios making it less dependable for precise predictions, than more sophisticated machine learning models. Wood's Model demonstrates comparable performance to these machine learning algorithms, indicating its effectiveness in modelling the relationship between variables and predicting outcomes. While Wood's Model may offer interpretability and insight into structural restrictions, the results in this context suggest that ML algorithms can provide competitive predictive accuracy. These findings highlight the versatility and effectiveness of both ML and conventional LC approaches in addressing predictive modelling tasks.

4.8 Performance Comparison of six Machine Learning Models and two Conventional Lactation Curve models across different dairy animals using STDMY

Figure (4.26) highlights the comparison of the efficiency of six machine learning methods over two conventional LC models like Gamma incomplete function and the polynomial regression function. The higher adjusted R^2 values obtained from ML models demonstrate their ability to better capture and model the complex relationships present in the MY data, resulting in more accurate predictions or estimations.

In Figure (4.26), various machine learning models were evaluated for their performance on three breeds: Murrah, Nili Ravi, and crossbred cattle. Among these models, Machine learning linear regression (MLR), Random Forest (RF), and XGBoost (XGB) stand out as the top performers across the breeds, demonstrating consistently high accuracy and robustness in regression tasks.

Among the three different types of breeds studied, Machine learning linear regression (MLR) stood out as the best-performing model, in all three cases boasting impressive accuracy rates Murrah (Table 4.11), Nili Ravi (Table 4.12), and crossbred cattle (Table 4.13). This highlights its reliability when it comes to regression tasks.

Upon comparing the performance of various ML algorithms using RMSE and MAE metrics, MLR emerges as the top performer with the lowest RMSE of 114.7233 and MAE of 79.1808, indicating its superior ability to minimize errors between predicted and actual values. Following MLR closely are eXtreme gradient boosting (XGB) and Random Forest (RF) with relatively lower error values compared to other algorithms. In contrast, K-Nearest Neighbors (KNN) exhibits the highest RMSE and MAE values, suggesting poorer performance in error minimization. These results underscore the importance of algorithm selection in achieving accurate predictions, with MLR demonstrating its effectiveness in this particular dataset.

Furthermore, both RF and XGB also showed competitive performance across the different dairy breeds. RF and XGB achieved adjusted R^2 accuracies for Murrah (Table 4.11), Nili Ravi (Table 4.12), crossbred cattle (Table 4.13). These ensemble methods are adept at capturing data relationships, which contributes to their effectiveness in regression tasks involving breeds.

These methods are excelling due to their traits and capabilities in dealing with regression tasks. Although MLR is typically used for regression purposes in this scenario, its directly employed as a regression model. MLR fits an equation to the observed data to minimize the variance between observed and predicted values. RF is a learning approach that merges decision trees to forecast continuous results. Each decision tree in the forest contributes to the prediction and the average or another form of aggregation of tree predictions is considered as the ultimate result. XGB is also a technique that constructs decision trees iteratively. In regression assignments, XGB forms a sequence of decision trees where each subsequent tree corrects its predecessor's errors ultimately leading to a regression model. The combined capability of these techniques to capture relationships within data, manage nonlinearities, and enhance performance plays a key role in their effectiveness at accurately forecasting animal attributes within a regression framework.

This underscores the advantage of employing machine learning approaches in such predictive tasks, especially when dealing with intricate datasets like STDMY as well as RTDMY records.

4.9 Estimation of genetic parameters for actual and predicted 305DMY

The 305DMY derived from the daily milk yield of animals along with the 305DMY predicted by the two conventional lactation curves (Wood's model and PRF) and the six ML (KNN, ANN, RF, SMOreg, MLR, XGB) models were arranged in a data file for all the animals. The pedigree of all the Murrah, Nili Ravi buffaloes and crossbred cattle under the study were prepared for the last three generations from the available records for estimation of the variance components and breeding values.

The 305DMY data was modelled in a repeatability animal model, taking period of calving, season of calving, and parity as fixed effects and animal and Permanent environment as the random effects. A combination of EM-REML for the first 200 rounds and thereafter AI-REML was used until convergence criteria of 10^{-14} were reached. The REML successfully converged for all three herds (Murrah, Nili Ravi buffaloes, and crossbred cattle) and the variance components were used for the estimation of heritability and repeatability for the 305DMY for the three herds. The standard deviations for the heritability and repeatability estimates were also calculated from the same software.

The purpose of the estimation of variance components, heritability, and repeatability was to compare the amount of genetic variance retained in the predicted 305DMY, as against the observed 305DMY, to compare the inflation or depletion of the genetic variability while prediction of 305DMY by different methods. Since the study intends to identify the 305DMY prediction method that can be used for further genetic evaluation of the animals, the comparison of the methods for their ability to retain the actual amount of genetic variability and actual genetic merit is crucial.

4.9.1 Heritability estimates in Murrah buffaloes

The heritability for the observed 305DMY was calculated as 0.2203±0.13229, indicating a moderate genetic influence on milk yield. This estimate of the heritability was on a little lower side, however in the range of reported values. The reason may have been the inclusion of 305DMY from the main as well as the segregated herd and little data exclusion as described in section 4.2, Table 4.4 (data standardization and normalization). However, the inclusion of a lot of variability in the 305DMY data was intended and deliberate to ensure that the trained models can accommodate a wide variation in lactation types, which however, led to smaller estimates for heritability with large standard deviations.

This estimate of heritability for 305DMY in Murrah buffaloes aligns with findings from Singh et al. (2015), who reported a low heritability estimate of 0.18±0.08 for FL305DMY, 0.17± 0.04 by Jamuna et al. (2015), and 0.19 reported by Tonhati et al. (2008) for Murrah buffaloes. Further for the Murrah breed, studies by Sigdel et al. (2015), Ramos et al. (2006), Rodrigues et al. (2010), Malhado et al. (2013) and Kumar et al. (2022) have also reported heritabilities of 0.18±0.14, 0.21, 0.25, 0.28 and 0.39±0.10 respectively. The study by Jakhar et al. (2016) revealed high very high heritability estimate for 305DMY as for 305DMY, estimated at 0.507±0.087. Earlier studies by Tailor et al. (1998), Dass and Sadana (2003), and Kumar (2000), also observed higher heritabilities in buffaloes.

The heritability estimates for predicted 305DMY in Murrah buffalo (Table 4.16) varied across different predictive modeling approaches. The Gamma Incomplete function shows a near zero heritability of 0.0082±0.0007. The Polynomial Regression Function also yielded a low heritability of 0.1414±0.1200, however, it was somewhat higher than Wood's model. The lower heritabilities are clearly indicating depletion of additive genetic variance estimated from Wood's model and PRF model approach.

Table 4.16: Heritability estimates for observed and predicted 305DMY

Methods of 305DMY Prediction	Murrah	Nili Ravi	CB Cattle
Observed	0.2203±0.1323	0.1794 ±0.2006	0.1852±0.1187
Wood's Model	0.0082±0.0007	0.0480±0.0000	0.1694±0.1195
PRF	0.1414±0.1200	0.0654± 0.0683	0.0657±0.0621
KNN	0.1815±0.1347	0.0391± 0.1936	0.1763±0.1105
ANN	0.1676±0.1679	0.0900± 0.1751	0.1203±0.1075
RF	0.1643±0.1321	0.1960±0.1962	0.1781±0.1160
SMOreg	0.1591±0.1324	0.1523±0.1933	0.1607±0.1137
MLR	0.1804±0.1307	0.1699±0.1941	0.1515±0.1123
XGB	0.1699±0.1317	0.1837± 0.1984	0.1740±0.1170

Among the ML models, the K Nearest Neighbor, Artificial Neural Network, Random Forest, Sequential Minimal Optimization Regression, Linear Model, and XGradient Boost models displayed varying heritability estimates, ranging from 0.1591 to 0.1815, which in turn were a little lower but closer to the heritability estimates from

the observed 305DMY, indicating a well-preserved variance components in the predicted 305DMY values. The consistency of the heritability estimates across different ML prediction methods indicates their possible usability for genetic evaluation for milk production traits in buffalo.

4.9.2 Heritability estimates in Nili Ravi buffaloes

The observed heritability for 305DMY in Nili Ravi buffaloes (Table 4.16) was found to be 0.2327±0.1876, very similar to that of Murrah buffaloes. Tamboli et al. (2021) and Suhail et al. (2009) reported a higher heritability estimate of 0.365±0.087 and 0.34 respectively for 305DMY, while Singh and Yadav (1987) and Bashir et al. (2017) found lower estimates as 0.13±0.05 and 0.105±0.02 in Nili Ravi buffaloes.

In comparison to the rest of the models, the heritability estimates for the predicted 305DMY by Wood's models was 0.0480±0.0000, indicating depletion of genetic variance during the prediction of 305DMY, an occurrence similar to the findings in Murrah buffaloes. Substantially reduced heritability was also observed for PRF (0.0654± 0.0683), rendering the usability of predicted 305DMY by standard lactation curve methods for genetic evaluation questionable.

Among the ML methods, The KNN model exhibits a lower heritability of 0.1314±0.1477, suggesting a depleted additive variance compared to the observed heritability. Conversely, the RF (0.2414±0.1872), SMOreg (0.2206±0.1776), MLR (0.2310±0.1840), and XGB (0.2531±0.1937) models demonstrate varying degrees of heritability, very close to the ones from the observed 305DMY. This comparison emphasizes the diversity of heritability estimates across different models, underlining the importance of selecting appropriate methodologies for accurate genetic evaluation in Nili Ravi buffaloes.

4.9.3 Heritability estimates in crossbred cattle

The heritability estimates for 305DMY in the studied population reveal variations across different models (Table 4.16) in crossbred cattle. The heritability estimate from observed 305DMY for crossbred cattle was found as 0.1852±0.1187.In crossbred cattle, Abou-Bakr et al. (2006) reported relatively low heritability estimates of 0.130 for 305DMY. Salem & Hammoud (2016) also found similar heritability estimates of 0.149±0.045 for the same trait. Salem et al. (2006) presented a moderate heritability estimate of 0.27 for 305DMY; whereas, El-Bayoumi et al. (2015) and

Radwan et al. (2015) reported slightly higher heritability estimates of 0.32 and 0.31 respectively for Holstein cows in Egypt.

Comparatively, for the estimated 305DMY, Wood's model yielded a slightly lower heritability of 0.1694±0.1195, while the PRF yielded the lowest heritability of 0.0657±0.0621.

The KNN model demonstrates a heritability of 0.1763±0.1105, indicating a level of genetic influence similar to the observed value. While the RF model exhibits a comparable heritability of 0.1781±0.11593. The SMOreg, MLR, and XGB models showed varying degrees of heritability as 0.1607±0.1137, 0.1515±0.1123, and 0.1740±0.1170, respectively; which were a little lower than the heritability estimates obtained from observed 305DMY.

On examining the estimates of variance components about heritability across different ML models and the conventional lactation curve model, it becomes evident that ML methodologies including MLR, RF, XGB, and ANN consistently exhibit superior and similar performances compared to other models across various breeds.

Through comprehensive analysis, these ML models consistently yield higher accuracy and predictive power in estimating heritability, as evidenced by observed values. This suggests their potential to enhance breeding strategies and genetic selection processes in livestock management. A marginal loss of additive genetic variance during the prediction process was as expected, and it should not pose a greater impact if the EBV estimates from the predicted 305DMY closely align with the EBV from the observed 305DMY.

4.9.4 Repeatability estimates

In Murrah buffaloes (Table 4.17), the repeatability estimates for observed 305DMY were found to be 0.4531±0.05535. Jamuna et al. (2015) and Sarkar et al. (2006) determined a lower repeatability estimate of 0.27±0.04 and 0.12 to 0.38 for 305DMY in Murrah buffaloes respectively.

Wood's model shows a notably higher repeatability of 1.2125±0.01397, indicating a substantial impact on the repeatability estimate. The PRF model exhibits a closer repeatability value of 0.4773±0.05776 compared to the observed value.

On the other hand, models like ANN (0.4274±0.05681) showed slightly lower and MLR (0.4573±0.05488), RF (0.4626±0.05457), SMOreg (0.4649±0.05461), XGB

(0.4605±0.05451) and KNN (0.4819±0.05370) showed repeatability values that are comparable or closer to the observed repeatability values in Murrah buffaloes.

Table 4.17: Repeatability estimates for observed and predicted 305DMY

Methods of 305DMY Prediction	Murrah	Nili Ravi	CB Cattle
Observed	0.4531±0.0553	0.3142±0.0946	0.3252±0.0981
Wood's Model	1.2125±0.0139	1.0000±0.0000	0.3303±0.0896
PRF	0.4773±0.0578	0.0654±0.0682	0.0657±0.0621
KNN	0.4819±0.0537	0.3354±0.0937	0.2746±0.0987
ANN	0.4274±0.0568	0.1421± 0.0987	0.2447±0.0980
RF	0.4626±0.0546	0.2915±0.0958	0.2882±0.0990
SMOreg	0.4649±0.0546	0.3218±0.0943	0.3201±0.0961
MLR	0.4573±0.0549	0.2848±0.0957	0.2810±0.0972
XGB	0.4605±0.0545	0.3142±0.0944	0.2915±0.0996

For Nili Ravi buffaloes (Table 4.17), the repeatability estimate for observed was found as 0.3417±0.1047. Sajjad Khan et al. (2007), Babar et al. (1996), Saxena and Tomar (1988) and Bashir et al. (2017) found the repeatability estimate for 305DMY to be a similar range at 0.46±0.02, 0.30±0.054, 0.321±0.032 and 0.35 and higher was reported by Syed et al. (1996) as 0.52±0.07 in Nili Ravi buffaloes. In Egyptian buffaloes, El-Itriby and Asker (1956) reported a repeatability estimate of 0.36, while Asker et al. (1963) found it to be 0.39. Juma and Al-Samarai (1985) obtained a somewhat higher estimate of 0.40.

Wood's model indicates a perfect repeatability of 1.0000±0.000, which may suggest a potential limitation or issue with the model in capturing the true repeatability. The Polynomial Regression Function model shows a repeatability of 0, indicating a potential discrepancy between the model and the observed data. Models like KNN (0.3075±0.0994), ANN (0.1741±0.1072), RF (0.3129±0.1073), SMOreg (0.3354±0.1032), MLR (0.3083±0.1062), and XGB (0.3349±0.1066) generally exhibit repeatability values that are lower than the observed repeatability in Nili Ravi buffaloes.

In crossbred cattle (Table 4.17), the observed repeatability is 0.3252±0.0981. Wood's model shows a repeatability of 0.3303±0.0896, which was consistent with the

observed value. The PRF indicates a repeatability of 0.0657±0.0621, suggesting a potential limitation of the model in capturing the true repeatability in crossbred cattle. Models like KNN (0.2746±0.0987), ANN (0.2447 ± 0.098), RF (0.2882±0.099), SMOreg (0.3201±0.0961), MLR (0.2810±0.0972), and XGB (0.2915±0.0996) generally display repeatability values that are comparable or slightly higher than the observed repeatability in crossbred cattle.

4.10 Estimated Breeding Values for observed and predicted 305DMY

The variance components estimated from REML were thereafter utilized to calculate the estimated breeding values (EBV) for all the animals in the pedigree using BLUP for the three breeds under study.

The purpose of the estimation and comparison of EBV was to check the amount of genetic superiority information retained in the predicted 305DMY as against the observed 305DMY. Further, the closeness of Estimated Breeding Values (EBVs) and animal ranking evaluated through Pearson's and Spearman's correlation of the EBVs of the animals calculated from the observed and estimated 305DMY will serve as the final criteria for identification of the best predictive modeling method for prediction of 305DMY. Since the study intends to identify the 305DMY prediction method that can be used for further genetic evaluation of the animals, the comparison of the methods for their ability to retain the actual amount of genetic variability and actual genetic merit is crucial.

4.10.1 Correlation among EBV for observed and predicted 305DMY in Murrah buffaloes

The depicted heatmap of the correlation matrix (Figure 4.27) illustrates the correlation coefficients, between the Estimated Breeding Values (EBV) from observed 305DMY and predicted 305DMY by models (Woods, Polynomial, KNN, ANN, RF, SMO, MLR, XGB) for Murrah buffaloes. These correlations show how closely the actual breeding values align with the predicted ones. The Pearson's correlation coefficients are shown on the upper diagonal of the matrix while the Spearman's rank correlations are displayed on the lower diagonal. Notably, strong positive correlations are observed across all models, reflecting their overall ability to capture and predict the breeding values effectively.

Upon examining the obtained correlations, it was observed that the relationships between EBV from observed 305DMY and EBV from predicted

305DMY using RF (0.99) XGB (0.99) and MLR (0.99) were near-perfect positive for both Pearson's and Spearman's correlation; suggesting very high similarity between them. In contrast, Wood's, ANN, and SMO exhibit negligible correlations. The correlation between EBV from the observed and PRF shows a correlation of 0.72, which is high but is nowhere near perfect correlation.

4.10.2 Correlation among EBV for observed and predicted 305DMY in Nili Ravi buffaloes

The heatmap of correlation matrix presented in Figure 4.28 displays the correlation coefficients between the EBV derived from observed 305DMY and predicted 305DMY by various models (Wood's, PRF, KNN, ANN, RF, SMO, MLR, and XGB) for Nili Ravi buffaloes.

Strong positive Pearson's correlations are evident between the EBV derived from observed 305DMY and predicted 305DMY by RF (0.993), MLR (0.99), KNN (0.971), and ANN (0.972). The Spearman's rank correlations for EBV also demonstrated a similar trend, where EBV derived from observed 305DMY and predicted 305DMY by RF (0.992), MLR (0.989), KNN (0.972), and ANN (0.966) models. These high correlation values suggest a significant similarity between the EBV obtained from observed 305DMY and predicted 305DMY of the ML models. The PRF model showed a lower correlation of 0.287 and 0.259 indicating a reasonable match while Wood's, SMO, and XGB exhibit weaker or negligible correlations in Pearson and Spearman correlation respectively.

4.10.3 Correlation among EBV for observed and predicted 305DMY in crossbred cattle

The heatmap of correlation matrix elucidates the relationships between Estimated Breeding Values (EBV) derived from observed and predicted 305DMY from various models in crossbred cattle (Figure: 4.29).

It was found evident that both the Pearson and Spearman correlations reveal robust associations among EBV for observed and predicted 305DMY, highlighting significant correlations by RF (0.99), SMO (0.99), MLR (0.98), XGB (0.99), KNN (0.98), ANN (0.96) and Wood's (0.93) models, whereas PRF exhibited 0.69 and 0.72 moderate associations in crossbred cattle.

Upon further analysis of correlation among EBV for observed and predicted 305DMY in Murrah, Nili Ravi buffaloes and crossbred cattle reveals distinct trends among models. MLR, RF, XGB, and KNN consistently showed near-perfect correlations with EBV obtained from observed 305DMY, indicating reliable predictive capabilities and highlighting their reliability in estimating breeding values, while the PRF model demonstrated slightly less correlation across all breeds.

4.11 Sire evaluation

Assessing sires is crucial for enhancing a breeding program, as it involves estimating the breeding values of bulls by analysing the production performance of their offspring, particularly daughters.

4.11.1 Correlation among sires' EBV for 305DMY in Murrah buffaloes

The EBV for the Murrah sires obtained from animal models was further used for generating Pearson's and Spearman's rank correlation. The spearman's rank correlation among the sires' EBV otherwise known as Sire Rank Correlation, has classically been a gold standard for comparison of methods for ranking the sires for their genetic merit since the sires contribute vastly to genetic improvement due to higher intensity of selection and large numbers of progenies. A near-perfect positive sire rank correlation ensures that the same sires will be selected by using either method and vice versa.

The heatmap of the correlation matrix, as shown in Figure 4.30, displays EBVs indicating the relationship between the EBV of sires from observed 305DMY and those predicted by different models (Wood's, PRF, KNN, ANN, RF, SMO, MLR, XGB) for Murrah sires. The sire rank correlations revealed that, all the EBV derived from predicted 305DMY using ML approaches viz. RF (0.991), MLR (0.99), ANN (0.984), SMO (0.984), KNN (0.983), and XGB (0.971), exhibited near-perfect sires rank correlation with the EBV derived from the observed 305DMY. Whereas PRF (0.713) displayed a positive sire rank correlation, it was largely away from perfect correlation and the ones from Wood's model (-0.048) showed negligible correlation.

4.11.2 Correlation among sires' EBV for 305DMY in Nili Ravi buffaloes

The heatmap of the correlation matrix presented in Figure 4.31 illustrates EBVs indicating the relationship between the EBV of sires from observed 305DMY and those predicted by different models (Wood's, PRF, KNN, ANN, RF, SMO, MLR,

XGB) for Nili Ravi sires. The Spearman correlations reveal strong positive relationships between EBVs obtained from observed 305DMY and predicted 305DMY by RF (0.99), XGB (0.98), MLR (0.95), SMO (0.97), KNN (0.93), and ANN (0.95) indicating high similarity. Conversely, the Wood's model (0.02) and PRF (-0.06) exhibit weak or negligible correlations.

4.11.3 Correlation among sires' EBV for 305DMY in crossbred cattle

The heatmap of the correlation matrix presented in Figure 4.32 represents EBVs indicating the relationship between the EBV of sires from observed 305DMY and those predicted by different models (Wood's, PRF, KNN, ANN, RF, SMO, MLR, XGB) for crossbred sires. The Spearman correlations reveal notable positive associations between EBVs obtained from observed 305DMY and predicted 305DMY by KNN (0.98), RF (0.98), SMO (0.98), XGB (0.971), MLR (0.961), ANN (0.942), and Woods (0.94) whereas PRF showed a moderate correlation of 0.643.

Overall, in assessing sire evaluation through Pearson's and Spearman's correlation coefficients between various ML models and the conventional LC model, it is evident that MLR, RF, XGB, and KNN consistently outperform other models across all breeds. In contrast, traditional lactation curve models often displayed lower correlation coefficients, indicating their limitations in accurately capturing the complexity of sire performance data. Therefore, the outperformance of ML-based approaches underscores their potential to revolutionize breeding strategies by offering more accurate and timely insights into sire potential, thereby driving genetic improvement and productivity in livestock populations.

4.12 Ranking for identification of Best Predictive Method for 305DMY

The four areas of model suitability for their evaluation were predictive performance for 305DMY using STDMY and RTDMY, preservation of variance components indicated by heritability and repeatability, and preservation of genetic merit indicated by rank correlation between EBVs from the observed and predicted 305DMY. To ensure no model with weak performance in any of the four criteria is retained, the worst performing model from each category was disqualified first-hand.

The models were evaluated for their predictive performance to predict 305DMY using the STDMY and RTDMY, with criteria of having lowest RMSE and MAE values. The closeness of heritability and repeatability estimates from predicted 305DMY with

that estimated from the observed 305DMY was the criteria for citing the model best as preserving the variance components in the prediction. Similarly, the highest animal and sire rank correlations between the EBVs from the observed and predicted 305DMY was used as the criteria for ranking the best model for genetic evaluation uses. The average of the ranks across the categories were used for final ranking of the models for their overall suitability.

4.12.1 Ranking for identification of Best Predictive Method for 305DMY in Murrah buffaloes

A comprehensive evaluation of various predictive methods for estimating 305DMY in Murrah buffaloes Table 4.18. Among these methods, (MLR) displayed superior predictive performance, as evidenced by its ability to minimize RMSE and MAE values, particularly when utilizing (STDMY) and (RTDMY). Following closely behind MLR in terms of predictive accuracy were XGBoost (XGB), Random Forest (RF), and Sequential Minimal Optimization Regression (SMOreg). Additionally, when considering genetic parameters, K-Nearest Neighbors (KNN) demonstrated closest proximity between predicted and observed values, with MLR and XGB also exhibiting strong predictive capabilities. Moreover, in terms of Estimated Breeding Value (EBV) correlation, RF and XGB emerged as particularly reliable methods, closely followed by MLR. Thus, based on these multifaceted evaluations, it is evident that both MLR and XGB deserve the top rank for their consistent and accurate prediction of 305DMY in Murrah buffaloes, with RF securing a commendable position following closely behind.

4.12.2 Ranking for identification of Best Predictive Method for 305DMY in Nili Ravi buffaloes

Upon considering Table 4.19 of various predictive methods for estimating 305DMY in Nili Ravi buffaloes, several criteria were employed to determine their efficacy. In the assessment using STDMY, MLR emerged as the top performer, boasting the lowest RMSE and MAE values, followed closely by XGBoost (XGB), Random Forest (RF), and Sequential Minimal Optimization Regression (SMOreg). Conversely, when considering RTDMY, XGB exhibited superior predictive capabilities, securing the first position, trailed by RF and MLR. However, when evaluating based on genetic parameters, RF and XGB demonstrated remarkable proximity between predicted and observed values, with SMOreg and MLR following.

Moreover, in terms of Estimated Breeding Value (EBV) correlation, RF and XGB maintained their prominence, with MLR also showing notable performance. Consequently, considering these diverse evaluations, it is evident that XGB and RF deserve the top rank for their consistent and accurate prediction of 305DMY in Nili Ravi cows, with MLR securing a commendable position following closely behind.

4.12.3 Ranking for identification of Best Predictive Method for 305DMY in crossbred cattle

In case of crossbred cows, after an evaluation of various predictive methods for estimating 305DMY Table 4.20. In the analysis using STDMY and RTDMY, MLR demonstrated superior predictive performance, showcasing the lowest RMSE and MAE values. MLR was closely followed by XGBoost (XGB) in terms of predictive accuracy. However, when assessing based on genetic parameters, Random Forest (RF) and Wood's method exhibited close proximity between predicted and observed values, with XGB also displaying noteworthy performance. Additionally, in terms of EBV correlation, Sequential Minimal Optimization (SMO) and XGB emerged as prominent predictors, with MLR and RF also performing well. Consequently, considering these various criteria, it is evident that RF deserves the top rank for its consistent and accurate prediction of 305DMY in crossbred cows, followed by XGB in second position, and MLR securing a commendable third place.

KNN, PRF, and Woods' method were disqualified due to their inferior predictive performance, misalignment with any of the four areas, contrasting with the superior predictive capabilities of MLR, XGB, and RF for predicting 305DMY across all breeds.

Table 4.18: Ranking for identification of Best Predictive Method for 305DMY in Murrah buffaloes

Parameters	STDMY			RTDMY			Genetic Parameter			EBV correlation			Overall ranking
	RMSE	MAE	Rank	RMSE	MAE	Rank	Heritability	Repeatability	Rank	Animal Model	Sire Model	Rank	
Wood's	189.4404	134.4258	VI	661.6894	520.6894	VI	0.0082±0.0007	1.2125±0.0139	NA	0	-0.0802	NA	NA
PRF	581.2491	223.6646	VII	14714.155	13827.919	NA	0.1414±0.1200	0.4773±0.0578	IV	0.8448	0.8471	VI	NA
KNN	1937.3052	1851.8578	NA	1937.0689	1851.7705	VII	0.1815±0.1347	0.4819±0.0537	I	0.9940	0.9748	III	NA
ANN	172.0332	135.1378	V	191.4103	154.0701	V	0.1676±0.1679	0.4274±0.0568	VII	0	0.9756	VI	V
RF	175.7308	118.8621	III	166.9495	118.9219	III	0.1643±0.1321	0.4626±0.0546	IV	0.9929	0.9899	I	III
SMOreg	202.0101	112.19	III	188.7394	115.5491	III	0.1591±0.1324	0.4649±0.0546	IV	0.0114	0.9855	V	VI
MLR	114.7233	79.1808	I	133.1428	101.6901	I	0.1804±0.1307	0.4573±0.0549	II	0.9884	0.9850	IV	I
XGB	159.3133	118.7779	II	141.0897	106.7128	II	0.1699±0.1317	0.4605±0.0545	II	0.9922	0.9911	I	I

NA: No ranking was allotted, as the model was worst performing among all models considered

Table 4.19: Ranking for identification of Best Predictive Method for 305DMY in Nili Ravi buffaloes

Parameters	STDMY			RTDMY			Genetic Parameter			EBV correlation			Overall ranking
	RMSE	MAE	Rank	RMSE	MAE	Rank	Heritability	Repeatability	Rank	Animal Model	Sire Model	Rank	
Wood's	173.6207	116.6779	V	727.9979	599.5509	VI	0.0480±0.0000	1.0000±0.0000	IV	0	0.0184	NA	NA
PRF	1204.0767	386.3577	VII	7048.09920	7744.1743	NA	0.0654± 0.0683	0.0654±0.0682	NA	0.2595	-0.0638	NA	NA
KNN	1642.653	1521.5132	NA	1642.3256	1521.2333	VII	0.0391± 0.1936	0.3354±0.0937	VI	0.9729	0.9289	IV	NA
ANN	211.8832	147.3982	VI	224.5380	169.0554	V	0.0900± 0.1751	0.1421± 0.0987	VII	0.9666	0.9481	IV	V
RF	160.0852	106.9202	III	151.2685	101.3406	II	0.1960±0.1962	0.2915±0.0958	I	0.9928	0.9876	I	I
SMOreg	176.2554	103.8218	IV	175.5199	104.1310	IV	0.1523±0.1933	0.3218±0.0943	III	-0.0328	0.9729	VI	IV
MLR	111.8572	75.1338	I	140.1911	104.5240	III	0.1699±0.1941	0.2848±0.0957	V	0.9896	0.9526	II	III
XGB	142.3964	99.9082	II	138.6985	102.5287	I	0.1837± 0.1984	0.3142±0.0944	I	0	0.9819	III	I

NA: No ranking was allotted, as the model was worst performing among all models considered

Table 4.20: Ranking for identification of Best Predictive Method for 305DMY in crossbred cattle

Parameters	STDMY			RTDMY			Genetic Parameter			EBV correlation			Overall ranking
	RMSE	MAE	Rank	RMSE	MAE	Rank	Heritability	Repeatability	Rank	Animal Model	Sire Model	Rank	
Wood's	372.819	248.134	II	1214.81	888.3409	VI	0.1694±0.1195	0.3303±0.0896	I	0.9386	0.9368	VII	NA
PRF	1561.632	597.9387	VII	17282.426	15213.034	NA	0.0657±0.0621	0.0657±0.0621	NA	0.7220	0.6431	NA	NA
KNN	3466.5978	3122.4017	NA	3466.4064	3122.1717	VII	0.1763±0.1105	0.2746±0.0987	V	0.9810	0.9778	III	NA
ANN	387.1738	298.4983	VI	449.7630	338.5875	V	0.1203±0.1075	0.2447±0.0980	VII	0.9634	0.9425	VI	V
RF	374.654	246.1364	II	438.0900	290.3945	III	0.1781±0.1160	0.2882±0.0990	I	0.9866	0.9771	II	I
SMOreg	532.252	233.4109	V	595.0341	271.5802	III	0.1607±0.1137	0.3201±0.0961	IV	0.9922	0.9828	I	IV
MLR	256.225	177.1694	I	367.6877	266.2413	I	0.1515±0.1123	0.2810±0.0972	VI	0.9778	0.9607	V	III
XGB	353.4422	258.973	II	403.2322	273.3606	II	0.1740±0.1170	0.2915±0.0996	III	0.9861	0.9710	IV	II

NA: No ranking was allotted, as the model was worst performing among all models considered

4.13 Societal Benefits

The current study offers significant societal benefits in dairy farming by enabling the utilization of random test day milk yield data for predicting 305DMY, thereby enhancing flexibility and accuracy in dairy herd management. This approach extends the forecasting horizon, providing farmers with insights into future milk production levels and allowing for better resource allocation. Furthermore, ML's flexibility in data recording encourages wider participation in Field Progeny Testing programs, facilitating the inclusion of a broader range of animals in genetic evaluations. This comprehensive coverage ensures that valuable genetic variation is captured, leading to more resilient and adaptable dairy herds. Additionally, with a larger number of animals included in Field Progeny Testing, ML enables more precise and intense selection of breeding bulls based on their genetic merit, ultimately improving milk yield performance and productivity in dairy operations.

CHAPTER V

SUMMARY AND CONCLUSION

Dairy production serves as a crucial sustenance factor and economic driver for rural communities in India, providing a vital source of livelihood for many. Therefore, the meticulous recording of milk yield stands as a significant asset for dairy farmers, aiding in effective farm management. Milk yield analysis serves as a cornerstone for genetic evaluation of dairy animals, playing a pivotal role in the selection and culling processes. Consequently, various traditional methods, including regression analysis, are employed to predict milk yield based on cow-related factors. However, these methods are hampered by inherent structural limitations, which hinder their effectiveness in providing accurate predictions.

The genetic evaluation of dairy animals heavily relies on the milk records of individual animals and their relatives. Conducting daily milk recordings is challenging in the Indian scenario due to scattered small and unorganized farms, time-consuming process, high cost, and other practical difficulties, especially in field conditions. As a result, the scope of genetic evaluation for the majority of animals is severely limited. Over time, several traditional techniques have been utilized to forecast the 305-days milk yield by leveraging test-day milk yield records. However, predictions from test day recording are prone to errors, particularly when deviations occur in the test day recording process. These challenges contribute to gaps in knowledge and hinder the effective genetic evaluation of dairy animals. So, an attempt to address this issue using ML models for prediction of 305DMY was tried and tested in the present study.

The data for this research was systematically gathered from records kept at the Directorate of Livestock Farms, Guru Angad Dev Veterinary and Animal Sciences University, located in Ludhiana, Punjab. This extensive dataset covers a period spanning from 2014 to 2022, totalling nine years. The study involved analysing 1662 lactation records from 940 dairy animals of three breeds (Murrah, Nili Ravi buffaloes and crossbred cattle). The 305-days milk yield (305DMY) of Murrah, Nili Ravi buffaloes, and crossbred cattle (CB cattle) was estimated by employing two traditional lactation curve methods and six different ML approaches, utilizing the standard test day milk yield records.

After pre-processing of the data, the dataset was parted into training, testing, and validation sets. Six machine learning models (KNN, ANN, RF, SMOreg, MLR, and XGB) were selected and trained for prediction of 305DMY. Initially, these models underwent training and testing using STDMY records and subsequently RTDMY records using the same machine learning algorithm.

The model efficiency parameters (R, R^2, adjusted R^2, MAE, MAPE, RMSE, RAE, and RRSE) for six ML algorithms, (KNN, ANN, RF, SMOreg, MLR, and XGB), and two conventional lactation curve methods (Gamma, and PRF), for predicting the 305DMY of dairy animals using STDMY and RTDMY were generated. Higher values of R, R^2, and adjusted R^2 and lower values of MAE and RMSE, indicate better model fit. Overall, MLR, RF, and XGB appear to be the most effective algorithms for predicting the 305DMY.

A novel method was utilized to generate a subset of a thousand RTDMY records for each breed, adhering to specific conditions. This approach ensured a systematic and controlled generation of random testing numbers, facilitating the evaluation and validation of the ML models with diverse and representative data points. It was observed that utilizing random test day milk yield records yielded better to comparable performance values compared to using standard test day milk yield records. This indicates that the models trained and tested with random test day milk yield data exhibited at-par or better predictive accuracy and generalization capabilities. While the conventional lactation curve such as Gamma Incomplete Function (Wood's model) demonstrated good performance when using STDMY records, it did not exhibit the same level of performance when utilizing RTDMY data. This discrepancy suggests that Wood's model may have limitations in capturing the variability present in the RTDMY records, leading to a reduction in predictive accuracy. The Polynomial Regression Function (PRF) model was found to be less accurate when using both STDMY records and RTDMY records. This indicates that PRF may not be well-suited for accurately predicting the 305DMY based on the available data and features, regardless of the type of test day milk yield data used. Overall, these findings highlight the importance of selecting appropriate machine learning algorithms.

The study estimated genetic parameters, including heritability and repeatability, for the observed and predicted 305DMY using a repeatability animal model, with period of calving, season of calving, and parity as fixed effects, along with animal and

permanent environment as random effects. This analysis aimed to compare the retention of genetic variance in predicted 305DMY against observed values. Upon examining the heritability across different ML models and conventional lactation curve models, it becomes evident that ML methodologies, including MLR, RF, and XGB consistently demonstrate superior performances across various breeds. A marginal loss of additive genetic variance during the prediction process was as expected, and it should not pose a significant impact if the estimated breeding values (EBVs) from the predicted 305DMY closely align with the EBVs from the observed 305DMY.

Overall, in assessing the animal and sire EBV obtained from the observed and predicted 305DMY using various ML models and the conventional LC model, it is evident that MLR, RF, XGB, and KNN consistently outperform other models across all breeds, by having near perfect positive correlation with EBV for observed 305DMY. In contrast, traditional lactation curve models often displayed lower correlation coefficients, indicating their limitations in accurately capturing the complexity of sire performance data, thus making them less suitable for predicting 305DMY for further genetic evaluation.

From the present study following conclusions can be drawn:

CONCLUSIONS

1. The random test day milk yield (RTDMY) can be as equally effective as STDMY in accurately predicting 305DMY, when using ML algorithms.

2. Among the machine learning methods studied, Machine learning Linear Regression (MLR) Random Forest (RF) and eXtreme Gradient Boosting (XGB) have emerged as particularly promising for estimating 305DMY due to their accuracy and robustness.

3. The predicted 305DMY values from machine learning models, such as MLR, XGB and RF can be used for estimating genetic parameters and Estimated Breeding Values (EBVs) for dairy animal breeding practices.

REFERENCES

Abdulazeez, A. M., Zeebaree, D., Abdulqader, D. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review Machine Learning View project How To Choose A Performance Metric View project Machine Learning Supervised Algorithms of Gene Selection: A Review.

Abou-Bakr, S., Alhammad, H. O. A., Sadek, R. R., & Nigm, A. A. (2006). Productive and reproductive characteristics of Holstein cows raised under intensive farming system in Egypt. *Egyptian Journal of Animal Production*, *43*(2), 91-98.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717-727.

Aguirre, D., & Boschini, C. (1992). Comparison of milk yield of indigenous and imported Jersey cows during three completed lactations. 4. Analysis of R 2 values for various models of the lactation curve. *Ciencias Veterinarias (Heredia), 14*(2), 15-23.

Akgün, B., & Öğüdücü, Ş. G. (2015, August). Streaming linear regression on spark MLlib and MOA. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 1244-1247).

Ali, T. E., & Schaeffer, L. R. (1987). Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, *67*(3), 637-644.

Al-Samarai, F. R., & Al-Kazaz, A. A. (2015). Applications of molecular markers in animal breeding. American Journal of Applied Scientific Research, *1(1)*, 1-5.

Angeles-Hernandez, J. C., Aranda-Aguirre, E., Muñoz-Benítez, A. L., Chay-Canul, A. J., Albarran-Portillo, B., Pollott, G. E., & Gonzalez-Ronquillo, M. (2021). Physiology of milk production and modelling of the lactation curve. *CABI Reviews*, (2021). https://doi.org/10.1079/PAVSNNR202116056

Asker, A. A., Bedeir, L. H., El-itriby, A. A., & Ahmed, I. A. (1963). Some genetic studies on the buffaloes in the UAR. *Genetics today. Proceedings of the XI International Congress of Genetics, The Hague, the Netherlands, 1,* 270

Aspilcueta-Borquis, R. R., Neto, F. A., Baldi, F., Bignardi, A. B., Albuquerque, L. G. D., & Tonhati, H. (2010). Genetic parameters for buffalo milk yield and milk quality traits using Bayesian inference. *Journal of Dairy Science*, *93*(5), 2195-2201.

Atıl, H., & Akıllı, A. (2016). Comparison of artificial neural network and K-means for clustering dairy cattle. *International Journal of Sustainable Agricultural Management and Informatics*, *2*(1), 40-52.

Aziz, M. A., Shalaby, N. A., El-Shafie, O. M., Mahdy, A. T., & Nishida, A. (2006). Comparison between the shapes of lactation curve of Egyptian buffalo milk yield estimated by the incomplete gamma function and a new model. *Livestock Research for Rural Development, 18*(5), 2006.

Babar, M. E., Yaqub, M., & Ahmad, T. (1996). Repeatability estimates of some performance characteristics in Nili-Ravi buffaloes. *Pakistan Journal of Agricultural Sciences, 85*, 87.

Babar, M. E., Yaqub, M., & Ahmad, T. (1996). Repeatability estimates of some performance characteristics in Nili-Ravi buffaloes. *Pakistan Journal of Agricultural Sciences*, *85*, 87.

BAHS. (2023). Basic animal husbandry and fisheries statistics. Department of Animal Husbandry, Dairying & Fisheries, Ministry of Agriculture, Government of India, New Delhi, India

Bajetha, G. 2006. Sire evaluation of Sahiwal and crossbred cattle. Ph.D. Thesis, G.B.Pant University of Agricultural & Technology, Pantnagar, India.

Bajetha, G., Singh, C. V., & Barwal, R. S. (2015). Sire evaluation on the basis of first lactation traits using best linear unbiased prediction (BLUP) method in sahiwal and crossbred cattle. *Livestock Research International*, *3*(4), 85-88.

Bajwa, I. R., Khan, M. S., Khan, M. A., & Gondal, K. Z. (2004). Environmental factors affecting milk yield and lactation length in Sahiwal cattle. *Pakistan Veterinary Journal*, *24*(1), 23-27.

Bangar, Y. C., & Verma, M. R. (2017). Non-linear modelling to describe lactation curve in Gir crossbred cows. *Journal of Animal Science and Technology*, *59*, 1-7.

Bangar, Y. C., Magotra, A., Malik, B. S., Malik, Z. S., & Yadav, A. S. (2021). Evaluating advanced computing techniques for predicting breeding values in Harnali sheep. *Tropical Animal Health and Production*, *53*, 1-6.

Banik S and Gandhi R S. 2010. Animal model versus conventional methods of sire evaluation in Sahiwal cattle. *Asian Australisian Journal of Animal Sciences 19*(9): 1225–28.

Banik, S. 2004. Sire evaluation in Sahiwal cattle. Ph.D. Thesis, NDRI (Deemed University), Karnal, Haryana, India.

Banu, R., Singh, A., Malhotra, R., Gowane, G., Kumar, V., Jaggi, S., & Raja, T. V. (2012). Comparison of different lactation curve models in Karan Fries cattle of India. *The Indian Journal of Animal Sciences*, *82*(11), 1377–1380.

Barbosa, S. B. P., Pereira, R. G. A., Santoro, K. R., Batista, A. M. V., & Neto, R. (2007). Lactation curve of cross-bred buffalo under two production systems in the Amazonian region of Brazil. *Italian Journal of Animal Science*, *6*(sup2), 1075-1078.

Barman, R. S. D., & Grewal, S. S. (2013). Effect of season on first lactation production performance of murrah buffaloes. *Indian Journal of Animal Production and Management*, *29*(1-2), 1-4.

Barman, R. S. D., Grewal, S. S., Dalal, D. S., Ahmad, M., & Mahto, D. (2012). Non-genetic factors affecting first lactation production traits in murrah buffaloes. *Indian Veterinary Journal*, *89*(8), 53-55.

Basant Singh, B. S., & Bhat, P. N. (1978). Some factors affecting the lactation curve in Hariana cattle. *Indian Journal of Animal Science*, *48*(10), 716–719.

Bashir, M. K., Khan, M. S., & Mustafa, M. I. (2017). Estimation of genetic parameters using animal model for some performance traits of Nili-Ravi buffaloes. *Pakistan Journal of Life & Social Sciences*, *15*(2), 120-125.

Bilal, G., & Khan, M. S. (2009). Use of test-day milk yield for genetic evaluation in dairy cattle: a review. *Pakistan Veterinary Journal*, *29*(1), 35-41.

Bilal, G., Khan, M. S., Bajwa, I. R., & Shafiq, M. (2009). Genetic control of test-day milk yield in Sahiwal cattle. *Pakistan Veterinary Journal*, *29*(1), 21-24.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32.

Brieuc, M. S., Waters, C. D., Drinan, D. P., & Naish, K. A. (2018). A practical introduction to Random Forest for genetic association studies in ecology and evolution. *Molecular ecology resources*, *18*(4), 755-766.

Brody, S., Ragsdale, A. C., & Turner, C. W. (1923). The rate of decline of milk secretion with the advance of the period of lactation. *The Journal of General Physiology*, *5*(4), 441-444.

Cankaya, S., Unalan, A., & Soydan, E. (2011). Selection of a mathematical model to describe the lactation curves of Jersey cattle. *Archives Animal Breeding*, *54*(1), 27-35.

Capuco, A. V., Ellis, S. E., Hale, S. A., Long, E., Erdman, R. A., Zhao, X., & Paape, M. J. (2003). Lactation persistency: insights from mammary cell proliferation studies. *Journal of animal science*, *81*(15_suppl_3), 18-31.

Catillo, G., Macciotta, N. P. P., Carretta, A., & Cappio-Borlino, A. (2002). Effects of age and calving season on lactation curves of milk production traits in Italian water buffaloes. *Journal of Dairy Science*, *85*(5), 1298-1306.

Chakraborty, D., Dhaka, S. S., Pander, B. L., Yadav, A. S., & Dandapat, A. (2010). Genetic studies on 305 days and test day milk yield records in Murrah buffaloes. *The Indian Journal of Animal Sciences*, *80*(8),729–732.

Cheema, J. S., & Basu, S. B. (1983). Usefulness of some components of the lactation curve for selection in Murrah buffaloes. *Animal Science*, *36*(2), 277-283.

Cilek, S., & Keskin, I. (2008). Comparison of six different mathematical models to the lactation curve of Simmental cows reared in Kazova state farm. *Journal of Animal and Veterinary Advances, 7*, 1316-1319.

Cobby, J. M., & Le Du, Y. L. P. (1978). On fitting curves to lactation data. *Animal Science*, *26*(2), 127-133.

Cobuci, J. A., Euclydes, R. F., Lopes, P. S., Costa, C. N., Torres, R. D. A., & Pereira, C. S. (2005). Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. *Genetics and Molecular Biology*, *28*, 75-83.

Coletta, A., Caso, C., Castrillo, M., Parlato, M., Zullo, A., & Zicarelli, L. (2007). Fit of the Wood function to milk yield data collected by different recording systems in Mediterranean Italian buffalo. *Italian Journal of Animal Science* 66 (Suppl.2): 503-505.

Cunha DNFV, Pereira JC, Silva FF, Campos OF, Braga JL, Martuscello JA. 2010. Selection of models of lactation curves to use in milk production simulation systems. *Rev Bras Zootecn.* 39:891–902.

Cunningham, E. P. (1965). The evaluation of sires from progeny test data. *Animal Science, 7*(2), 221-231.

Dahiya DS, Khana AS, Singh RP. Effectiveness of sire evaluation for milk yield with auxiliary traits in Hariana cattle. *Indian Journal of Animal Sciences* 75(5):518-523.

Dalal, D. S., Rathi, S. S., & l Raheja, K. (1999). Genetic evaluation of Hariana sires for test day milk records and first lactation milk yield. *The Indian Journal of Animal Sciences*, *69*(12),1042-1043.

Dass, G. 1995. Studies on prediction of lactation yield based on test-day values in buffaloes. Ph.D Thesis, NDRI (Deemed University), Karnal, Haryana, India.

Dass, G. and Sadana, D. K. (2003). Predictability of lactation milk yield based on test day values in Murrah buffaloes. *Indian Journal of Animal Research. 37(2):* 136-138.

Dave, B. K. (1971). First lactation curve of the Indian water buffalo. *JNKVV Research Journal, 5*, 93–98.

De Boer, J. A., Weller, J. I., Gipson, T. A., & Grossman, M. (1989). Multiphasic analysis of milk and fat yield curves of Israeli Holsteins. *Journal of Dairy Science*, *72*(8), 2143-2152.

Dehghan, M. H., Hamidi, F., & Salajegheh, M. (2015, September). Study of linear regression based on least squares and fuzzy least absolutes deviations and its application in geography. In *2015 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)* (pp. 1-6). IEEE.

Dematawewa, C. M. B., Pearson, R. E., & VanRaden, P. M. (2007). Modeling extended lactations of Holsteins. *Journal of Dairy Science*, *90*(8), 3924-3936.

Deulkar, P. B. and Kothekar, M. D. 1999. Sire evaluation considering first lactation yield for improvement of lifetime production in Sahiwal. *Indian Journal of Animal Sciences,69(4):* 240-242

Dhaiya, D. S., & Khanna, A. S. (2005). Effectiveness of sire evaluation for milk yield with auxiliary traits in Hariana cattle. *The Indian Journal of Animal Sciences*, c*75*(5), 518-523.

Dhaka, S. S., & Raheja, K. L. (2000). A comparison of sire evaluation methods. *The Indian Journal of Animal Sciences*, *70*(6), 643-644.

Dhirendra Kumar, D. K., Verma, S. B., Singh, S. R., Mandal, K. G., & Pandey, R. P. (2005). Effect of genetic and non-genetic factors on dry period and age at first

calving in Haryana and its crosses. *Indian Journal of Animal Health, 44*(2), 99-102.

Dijkstra, J., France, J., Dhanoa, M. S., Maas, J. A., Hanigan, M. D., Rook, A. J., & Beever, D. E. (1997). A model to describe growth patterns of the mammary gland during pregnancy and lactation. *Journal of Dairy Science*, *80*(10), 2340-2354.

Dijkstra, J., López, S., Bannink, A., Dhanoa, M. S., Kebreab, E., Odongo, N. E., & France, J. (2010). Evaluation of a mechanistic lactation model using cow, goat and sheep data. *The Journal of Agricultural Science*, *148*(3), 249-262.

Dimauro, C., Catillo, G., Bacciu, N., & Macciotta, N. P. P. (2005). Fit of different linear models to the lactation curve of Italian water buffalo. *Italian Journal of Animal Science*, *4*(sup2), 22-24.

Dimauro, C., Catillo, G., Bacciu, N., & Macciotta, N. P. P. (2005). Fit of different linear models to the lactation curve of Italian water buffalo. *Italian Journal of Animal Science*, *4*(sup2), 22-24.

Divya P, Singh A, Gandhi R S and Singh R K. 2014. Estimation of breeding values of first lactation 305-day milk yield from single and multi-trait animal models in Karan Fries cattle. *Indian Journal of Animal Sciences* 84: 1085–89.

Dohare, A. K., Singh, B., Verma, M. R., Perme, B., Sharma, V. B., Gupta, N., & Kshandakar, S. (2014). Comparison of standard lactation curve models using fortnightly milk records in Frieswal cattle. *Veterinary World*, *7*(10),831-834.

Dongre, V. B. (2012). *Modeling lactation curve for sire evaluation in Sahiwal cattle* (Doctoral dissertation, Ph. D. Thesis, NDRI, Karnal, India).

Dongre, V., Gandhi, R. S., & Singh, A. (2012). Comparison of different lactation curve models in Sahiwal cows. *Turkish Journal of Veterinary & Animal Sciences*, *36*(6), 723-726.

Dubey, P. P., Singh, C., & Barwal, R. S. (2014). Sire evaluation considering first lactation yield for improvement of life time production in Sahiwal and crossbred Cattle. *Advances in Animal and Veterinary Sciences*, *2*(1), 56-62.

Edwards, J. (1932). The Progeny Test as a Method of Evaluating the Dairy Sire. *The Journal of Agricultural Science*, *22*(4), 811-837.

El-Arian, M. N. (1986). *Genetic analysis of Murrah buffalo herds* (Doctoral dissertation, NDRI).

El-Bayoumi, K. M., El-Tarabany, M. S., Abdel-Hamid, T. M., & Mikaeil, O. M. (2015). Heritability, genetic correlation and breeding value for some productive and reproductive traits in Holstein cows. *Research Opinions in Animal and Veterinary Sciences, 5*(2), 65-70.

El-Itriby, A. A., & Asker, A. A. (1956). Repeatability and heritability of some dairy characters in cattle and buffaloes in Egypt. *Indian Journal of Dairy Science*, *9*, 157-163.

El-Tahawy, A. S., & El-Far, A. H. (2010). Influences of somatic cell count on milk composition and dairy farm profitability. *International Journal of Dairy Technology*, *63*(3), 463-469.

Erdem, H., Atasever, S., & Kul, E. (2010). Determination of milk production characteristics and milk losses related to somatic cell count in Jersey cows raised in the Black Sea region of Turkey. *Asian Journal of Animal and Veterinary Advances*, *5*(3), 217-222.

Erfani-Asl, Z., Hashemi, A., & Farhadian, M. (2015). Estimates of repeatability and heritability of productive trait in Holstein dairy cattle. *Iranian Journal of Applied Animal Science*, *5*(4), 827-832.

Ertuğrul, K. U. L., ABACI, S. H., Çobanoğlu, Ö., Gürcan, E. K., & Çankaya, S. (2018). Estimation of genetic and environmental parameters for milk traits in jersey cows. *Mediterranean Agricultural Sciences*, *31*(3), 311-315.

Espinosa AP, Rodríguez AF, Jiménez CJ, Espinoza VJL, Núñez DR. Genetic evaluation of a Holstein dairy herd in Baja California Sur, utilizing an animal model with repeated measurements. *Agrociencia, 35*(3), 347-353.

Falconer DS. Introduction to quantitative genetics. 3rd edition Longman, London, U.K, 1989, 438.

Fernández, C., Soria, E., Martin, J. D., & Serrano, A. J. (2006). Neural networks for animal science applications: Two case studies. *Expert Systems with Applications*, *31*(2), 444-450.

Fernández, C., Soria, E., Sanchez-Seiquer, P., Gómez-Chova, L., Magdalena, R., Martín-Guerrero, J. D., & Serrano, A. J. (2007). Weekly milk prediction on dairy goats using neural networks. *Neural Computing and Applications*, *16*, 373-381.

Ferreira, A. G., Henrique, D. S., Vieira, R. A., Maeda, E. M., & Valotto, A. A. (2015). Fitting mathematical models to lactation curves from holstein cows in the southwestern region of the state of Parana, Brazil. *Anais da Academia Brasileira de Ciências*, *87*, 503-517.

Ferreira, W. J., Teixeira, N. M., Euclydes, R. F., Verneque, R. D. S., Lopes, P. S., Torres, R. D. A., & Magalhães Júnior, M. N. (2003). Genetic evaluation of Holstein cattle using test day milk yield. *Revista Brasileira de Zootecnia*, *32*, 295-303.

Firat, M. Z. (2001). Sürü-y› l-ay Etkilerini fiansa Ba¤ l› Varsayan Dengesiz Kar› fl› k Bir Modelde Kontrol Günü Süt Verimlerinin Bayesian Analizi. *Turkish Journal of Veterinary & Animal Sciences,25*, 327-333.

Gahlot, G. C., Gahlot, R. S., & Jairath, L. K. (1988). Pattern of lactation curve in Rathi cattle. *Indian Journal of Animal Sciences,* 1988; 58(9): 1112– 1114.

Gandhi, R. S., & Gurnani, M. (1988). Optimum combinations of early traits for prediction of 8-year lifetime production in Sahiwal cattle. *Asian Journal of Dairy Research 7*, 7–12.

Gandhi, R. S., & Gurnani, M. (1991). Accuracy of different methods of sire evaluation in Sahiwal cattle. *Indian Veterinary Journal, 68*(7), 656-662

Geetha, E. (2005). *Studies on breeding values for persistency in Murrah buffaloes.* (M.Sc. Thesis, NDRI (Deemed University), Karnal, India).

Geetha, E., Chakravarty, A. K., & Kumar, K. V. (2006). Genetic persistency of first lactation milk yield estimated using random regression model for Indian Murrah buffaloes. *Asian-Australasian Journal of Animal Sciences*, *19*(12), 1696-1701.

Gianola, D., Okut, H., Weigel, K. A., & Rosa, G. J. (2011). Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC genetics*, *12*, 1-14.

Girimal, D., Kumar, D., Shahi, B. N., Ghosh, A. K., & Kumar, S. (2022). Sire evaluation using conventional methods and animal models in Sahiwal cattle. *Indian Journal of Animal Sciences*, *92*(4), 492-96.

Godara, A., Singh, D., & Grewal, S. S. (2004). Factors affectivity production, reproduction and lifetime traits and their association with production life. *Indian Buffaloes Journal*, *2*(1), 23-27.

Gore, S. S., Gour, A. K., Khothekar, M. D., & Deshkukh, S. N. (1996). A study on lactation curve in Sahiwal and Jersey x Sahiwal crossbreed cows. *PKV Research Journal, 20*(1), 47-49.

Gorgulu, O. (2012). Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South African Journal of Animal Science, 42*(3), 280-287.

Græsbøll, K., Kirkeby, C., Nielsen, S. S., Halasa, T., Toft, N., & Christiansen, L. E. (2016). Models to estimate lactation curves of milk yield and somatic cell count in dairy cows at the herd level for the use in simulations and predictive models. *Frontiers in Veterinary Science, 3*, 115.

Grossman, M., & Koops, W. J. (2003). Modeling extended lactation curves of dairy cattle: A biological basis for the multiphasic approach. *Journal of Dairy Science, 86*(3), 988-998.

Grzesiak, W., Błaszczyk, P., & Lacroix, R. (2006). Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture, 54*(2), 69-83.

Grzesiak, W., Lacroix, R., Wójcik, J., & Blaszczyk, P. (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science, 83*(2), 307-310.

Guevara, L., Castro-Espinoza, F., Fernandes, A. M., Benaouda, M., Muñoz-Benítez, A. L., del Razo-Rodríguez, O. E., ... & Angeles-Hernandez, J. C. (2023). Application of Machine Learning Algorithms to Describe the Characteristics of Dairy Sheep Lactation Curves. *Animals, 13*(17), 2772.

Guo, Z., & Swalve, H. (1995). Modelling of the lactation curve as a sub-model in the evaluation of test day records. *Interbull Bulletin*, (11).

Guo, Z., & Swalve, H. H. (1997). Comparison of different lactation curve sub-models in test day models. *Interbull Bulletin, 16*, 75-75.

Gupta, R., Sahoo, S. K., Kaur, S., Dash, S. K., Malhotra, P., & Bansal, B. K. (2020). Effect of non-genetic factors on test-day milk yields and first lactation milk yields in Holstein Friesian× Sahiwal Cattle. *Journal of Animal Research, 10*(3), 447-451.

Gurmessa, J. and Melaku, A. (2012). Effect of lactation stage, pregnancy, parity and age on yield and major components of raw milk in bred cross Holstein Friesian Cows. *World Journal of Dairy & Food Sciences* 7: 146-149.

Hagger, C., & Dempfle, L. (1982). What can be gained by the BLUP procedure under various circumstances?. In *Annales de génétique et de sélection animale* (Vol. 14, No. 1, pp. 122-123).

Hamadani, A., Ganai, N. A., Mudasir, S., Shanaz, S., Alam, S., & Hussain, I. (2022). Comparison of artificial intelligence algorithms and their ranking for the prediction of genetic merit in sheep. *Scientific Reports*, *12*(1), 18726.

Hammami, H., Rekik, B., Soyeurt, H., Gara, A. B., & Gengler, N. (2008). Genetic parameters for Tunisian Holsteins using a test-day random regression model. *Journal of Dairy Science*, *91*(5), 2118-2126.

Harvey W R. 1990. Guide for LSMLMW, PC-1 Version, mixed model least squares and maximum likelihood computer programme, January 1990. Mimeograph Ohio State University, USA.

Hempstalk, K., McParland, S., & Berry, D. P. (2015). Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. *Journal of Dairy Science*, *98*(8), 5262-5273.

Henderson, C. R. (1973). Sire evaluation and genetic trends. *Journal of Animal Science*, *1973*(Symposium), 10-41.

Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *Journal of Dairy Science*, *57*(8), 963-972.

Henderson, C. R. (1975). Use of relationships among sires to increase accuracy of sire evaluation. *Journal of Dairy Science*, *58*(11), 1731-1738.

Jadhav, R.G., Vaidya, M.S. & Toro, V.A., 1998. Prediction of first lactation milk yield based on part lactation performance in crossbred cows. *Journal of Bombay Veterinary College*, 6, 57-60.

Jain, A. 1996. Sire evaluation under multiple trait models in Murrah buffaloes. Ph. D. Thesis, NDRI (Deemed University), Karnal, Haryana, India.

Jain, A., & Taneja, V. K. (1982). Effects of non-genetic factors on reproduction and production traits in Murrah buffaloes. *Asian Journal of Dairy Research, 1*(2), 123-129.

Jakhar, V. I. K. R. A. M., Vinayak, A. K., & Singh, K. P. (2016). Genetic evaluation of performance attributes in Murrah buffaloes. *Haryana Veterinarian*, *55*(1), 66-69.

Jakhar, V., Yadav, A. S., & Dhaka, S. S. (2017). Estimation of genetic parameters for production and reproduction traits in Murrah buffaloes. *International Journal of Current Microbiology and Applied Sciences*, *6*(11), 4297-4303.

Jamrozik J and Schaeffer L R. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *Journal of Dairy Science* 80: 762–70.

Jamrozik, J., & Schaeffer, L. R. (1997). Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *Journal of Dairy Science*, *80*(4), 762-770.

Jamrozik, J., Jansen, G., Schaeffer, L. R., & Liu, Z. (1998). Analysis of persistency of lactation calculated from a random regression test day model. *Interbull Bulletin*, *17*, 110-110.

Jamrozik, J., Kistemaker, G. J., Dekkers, J. C. M., & Schaeffer, L. R. (1997). Comparison of possible covariates for use in a random regression model for analyses of test day yields. *Journal of Dairy Science*, *80*(10), 2550-2556.

Jamrozik, J., Schaeffer, L. R., & Dekkers, J. (1996). Random regression models for production traits in Canadian Holsteins. *Interbull Bulletin*, (14), 124-124.

Jamrozik, J., Schaeffer, L. R., & Dekkers, J. C. M. (1997). Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science*, *80*(6), 1217-1226.

Jamuna, V., Chakravarty, A. K., Singh, A., & Patil, C. S. (2015). Genetic parameters of fertility and production traits in Murrah buffaloes. *Indian Journal of Animal Research*, *49*(3), 288-291.

Japheth, K. P., Mehla, R. K., & Bhat, S. A. (2015). Effect of non- genetic factors on various economic traits in Karan Fries crossbred cattle. *Indian Journal of Dairy Science*, *68*(2), 163-169.

Jensen, J. (2001). Genetic evaluation of dairy cattle using test-day models. *Journal of Dairy Science*, *84*(12), 2803-2812.

Jingar, S. C., Mehla, R. K., Singh, M., & Singh, P. K. (2014). Effect of stages and level of milk production on mastitis incidence in cows and murrah buffaloes. *Journal of Bio Innovation*, *3*(3), 117-123.

Jingar, S., Mehla, R. K., Singh, M., & Roy, A. K. (2014). Lactation curve pattern and prediction of milk production performance in crossbred cows. *Journal of Veterinary Medicine*, *2014*.

Jónás, E. M., Atasever, S., Gráff, F. M., & Erdem, H. (2016). Non-genetic factors affecting milk yield, composition and somatic cell count in Hungarian Holstein cows. *Journal of the Faculty of Veterinary Medicine, Kafkas University*, *22*, 361-366.

Juma, K. H., & Al-Samarai, W. W. (1985). Some economic traits of Iraqi buffaloes. I. Dairy characteristics. *World Review of Animal Production, 21*, 67-70.

Kaliappan, J., Srinivasan, K., Mian Qaisar, S., Sundararajan, K., & Chang, C. Y. (2021). Performance evaluation of regression models for the prediction of the COVID-19 reproduction rate. *Frontiers in Public Health*, *9*, 729795.

Kannan, D., & Gandhi, R. S. (2006). Prediction of lifetime production in Sahiwal cattle. *Indian Journal of Animal Sciences*, *76*(9), 768.

Kasarda, R., Moravčíková, N., Mészáros, G., Simčič, M., & Zaborski, D. (2023). Classification of cattle breeds based on the random forest approach. *Livestock Science*, *267*, 105143.

Kathiravan, P. 2009. Genetic evaluation of lifetime performance of Sahiwal cattle. Ph.D. Thesis, NDRI (Deemed University), Karnal, Haryana, India.

Katneni, V. K. (2007) *Studies on genetic persistency of milk production in Murrah buffaloes*. (Ph.D. Thesis, NDRI, (Deemed University), Karnal, India).

Kaya, İ., Akbaş, Y., & Uzmay, C. A. N. (2003). Estimation of breeding values for dairy cattle using test-day milk yields. *Turkish Journal of Veterinary & Animal Sciences*, *27*(2), 459-464.

Kaygisiz, A. (2013). Estimation of genetic parameters and breeding values for dairy cattle using test-day milk yield records. *The Journal of Animal & Plant Sciences*, *23*(2): 345-349.

Kettunen, A., E. A. Mäntysaari, I. Strandén, J. Pösö, and M. Lidauer. 1998. Estimation of genetic parameters for first lactation milk production using random regression models. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia, 11-16 January, 1998. Volume 23: Opening: plenary; technology; transfer; beef cattle; dairy cattle; pigs., 1998, 307-310.

Khan, M. S., Hyder, A. U., Bajwa, I. R., Rehman, M. S., & Hassan, F. (2005). Prediction of lactation yield from last-record day and average daily yield in Nili-Ravi buffaloes. *Pakistan Veterinary Journal*, *25*, 175-178.

Khosla, S. K., Gill, S. S., & Malhotra, P. K. (1984). Effect of some non-genetic factors on age at calving and service period in herdbook-registered Murrah buffaloes under village conditions. *Indian Journal of Animal Science, 54*, 407-412.

Koçak, Ö. M. Ü. R., & Ekiz, B. Ü. L. E. N. T. (2008). Comparison of different lactation curve models in Holstein cows raised on a farm in the south-eastern Anatolia region. *Archives Animal Breeding*, *51*(4), 329-337.

Kokate, L. S., Singh, A., Banu, R., Gandhi, R. S., Chakravarty, A. K., Gupta, A. K., & Sachdeva, G. K. (2014). Prediction of 305-day lactation milk yield based on bimonthly test day values in Karan Fries cattle. *Indian Journal of Animal Research*, *48*(2), 103-105.

Kolte, D. V., Gore, A. K., & Deshmukh, S. N. (1986). Study of lactation curve in Sahiwal breed of cattle. *PKV Research Journal*, *10*(2), 145-147.

Kominakis, A. P., Abas, Z., Maltaris, I., & Rogdakis, E. (2002). A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computers and Electronics in Agriculture*, *35*(1), 35-48.

Kong, L. N., Li, J. B., Li, R. L., Zhao, X. X., Ma, Y. B., Sun, S. H., & Zhong, J. F. (2018). Estimation of 305-day milk yield from test-day records of Chinese Holstein cattle. *Journal of Applied Animal Research*, *46*(1), 791-797.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3-24.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1-26.

Kumar R., & Bhat P. N. (1979). Lactation curves in Indian buffaloes. *Indian Journal Dairy Science*, *32*(2), 156-160.

Kumar, A. (2000). *Genetic Studies on First Lactation Traits in Murrah Buffaloes* (Doctoral dissertation, Animal Breeding, CCSHAU, Hisar, India).

Kumar, A., Gupta, J. P., Kaswan, S., & Lathwal, S. S. (2012). Non-genetic factors influencing monthly test day milk yield records in Murrah buffaloes. *Indian Journal of Field Veterinarians (The)*, *8*(1), 41-43.

Kumar, A., Rana, Z. S., Dhaka, S. S., Dalal, D. S., Singh, S., & Sethi, R. K. (2002). Test day yield and lactation parameters of Murrah buffaloes. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002. Session 7* (pp. 0-3). Institut National de la Recherche Agronomique (INRA).

Kumar, A., Vohra, V., Verma, U., & Singh, U. (2022). Estimates of genetic parameters for production and reproduction traits in Murrah buffaloes (Riverine buffalo) in India. *Buffalo Bulletin*, *41*(1), 135-141.

Kumar, D., Singh, H., Kumar, D., & Singh, C. V. (2003). Non-genetic factors affecting some production and reproduction traits in Murrah buffaloes. *Indian Journal of Animal Research*, *37*(1), 24-27.

Kumar, R. (1992). *Effectiveness of additive and multiplicative age correction factor for sire evaluation in Murrah buffaloes.* (M.Sc. Thesis, NDRI (Deemed University), Karnal, India).

Kumar, R. and Gandhi, R. S. 1994. Additive versus multiplicative age correction factors for Murrah sire evaluation. *Buffalo Journal, 10*(3): 205-212.

Kumar, R., & Bhat, P. N. (1979). Lactation curve in Indian buffaloes. *Indian Journal of Dairy Science, 32*(2), 156-160.

Kumar, S. (2007). A comparative study of lactation curve models for prediction of lactation milk yield in Murrah buffaloes. (M.V.Sc. Thesis, National Dairy Research Institute (Deemed University), Karnal, India).

Kumar, S., Gupta, I. D., Sharma, N., Deginal, R., Kumar, A., Chauhan, A., & Verma, A. (2021). Effect of season, parity and stage of lactation on productive performance of Sahiwal cattle. *Indian Journal of Animal Research*, *55*(5), 597-602.

Kumar, V., Chakravarty, A. K., Magotra, A., Patil, C. S., & Shivahre, P. R. (2019). Comparative study of ANN and conventional methods in forecasting first lactation milk yield in Murrah buffalo. *Indian Journal of Animal Sciences*, *89*(11), 1262-68.

Kumar, V., Chakravarty, A. K., Patil, C. S., Valsalan, J., Sharma, R. K., & Mahajan, A. (2014). Genetic studies on first lactation monthly test day and 305 days milk yields in Murrah buffaloes. *Indian Veterinary Journal*, *91*(7), 31-33.

Kumari, R. (2019). Genetic variants in FAM19A1, KCNB1 and EDN3 gene and their association with udder type traits, clinical mastitis and milk production in karan fries and sahiwal cattle. Ph.D Thesis. ICAR-National Dairy Research Institute, Karnal, India.

Lacroix, R., Salehi, F., Yang, X., & Wade, K. (1997). Effects of data pre-processing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Transactions of the ASAE, 40*, 839–846.

Lacroix, R., Wade, K. M., Kok, R., & Hayes, J. F. (1995). Prediction of cow performance with a connectionist model. *Transactions of the ASAE*, *38*(5), 1573-1579.

Lalhmingmawii. (2021). *Study on abnormal lactation length in Murrah buffaloes.* (M.V.Sc. Thesis, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, India).

Ledic, I. L., Verneque, R. D. S., El Faro, L., Tonhati, H., Martinez, M. L., Oliveira, M. D. S. D., & Fernandes, L. D. O. (2002). Genetic evaluation of Gir breed sires for milk production at the control day and 305 days of lactation. *Revista Brasileira de Zootecnia*, *31*, 1964-1972.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Lim, H. I. (2019, July). A linear regression approach to modeling software characteristics for classifying similar software. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 942-943). IEEE.

Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., & Hostens, M. (2020). Leveraging latent representations for milk yield prediction and interpolation using deep learning. *Computers and Electronics in Agriculture*, *175*, 105600.

Liu, Z., Reents, R., Reinhardt, F., & Kuwan, K. (2000). Approaches to estimating daily yield from single milk testing schemes and use of am-pm records in test-day model genetic evaluation in dairy cattle. *Journal of Dairy Science, 83*(11), 2672-2682.

Livestock Census. (2012). 19th All India Livestock Census. Department of Animal Husbandry, Dairying & Fisheries. Ministry of Agriculture, Govt. of India.

Lodhi, G., Singh, C. V., Barwal, R. S., Shahi, B. N., & Dalal, D. S. (2016). Estimation of breeding values by different sire evaluation methods for selection of sires in crossbred cattle. *Journal of Veterinary Science & Medical Diagnosis*, *3*(10), 145-150

LR, S. (1994). Random regressions in animal models for test-day production in dairy cattle. In *World Congress of Genetics Applied Livestock Production, 1994* (Vol. 18, pp. 443-446).

Macciotta, N. P. P., Vicario, D., & Cappio-Borlino, A. (2005). Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science*, *88*(3), 1178-1191.

Macciotta, N. P., Dimauro, C., Rassu, S. P., Steri, R., & Pulina, G. (2011). The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, *10*(4), e51.

Macciotta, N., Vicario, D., Pulina, G., & Cappio-Borlino, A. (2002). Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *Journal of Dairy Science*, *85*, 3107-3114.

Malhado, C. H. M., Malhado, A. C. M., Ramos, A. D. A., Carneiro, P. L. S., Souza, J. C. D., & Pala, A. (2013). Genetic parameters for milk yield, lactation length and calving intervals of Murrah buffaloes from Brazil. *Revista Brasileira de Zootecnia*, *42*, 565-569.

Malhotra, P. (2014). *Genetic evaluation of murrah buffaloes at organised herds.* (Ph.D. Dissertation, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, India).

Mammadova, N., & Keskin, İ. (2013). Application of the support vector machine to predict subclinical mastitis in dairy cattle. *The Scientific World Journal*, e603897.

Masood, A. D., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine learning supervised algorithms of gene selection: *A review. Machine Learning*, *62*, 03.

Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, *1*(2), 140-147.

McQueen, R. J., Garner, S. R., Nevill-Manning, C. G., & Witten, I. H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, *12*(4), 275-293.

Mehta H. (2020). Study on abnormal lactation length in crossbred cattle. (M.V.Sc. Thesis, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, India).

Meyer, K. (1989). Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science*, *21*(2), 87-100.

Michalski, R. S., J. G. Carbonell, and T. M. Mitchell. 2013. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, Berlin, Germany.

Misztal, I., Duangjinda, M., & Tsuruta, S. (2004). BLUPF 90, Dairy Pack Version 2; Genetic Evaluation Program for Dairy Cattle. *Department of Animal and Dairy Science, The University of Georgia, Athens, GA30602*.

Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press.

Mrode, R. A., Swanson, G. J. T., & Lindberg, C. M. (2002). Efficiency of part lactation test day records for genetic evaluations using fixed and random regression models. *Animal Science*, *74*(2), 189-197.

Mukherjee, S. (2005). *Genetic evaluation of Frieswal cattle*. (Ph.D. Dissertation, NDRI (Deemed University), Karnal, India).

Mukherjee, S. 2005. Genetic evaluation of Frieswal cattle. Ph.D Thesis, NDRI (Deemed University), Karnal, Haryana, India.

Murphy, M., O'Mahony, M., Shalloo, L., French, P., & Upton, J. (2014). Com- parison of modelling techniques for milk-production forecasting. *Journal of Dairy Science*, *97*, 3352 – 3363.

Murphy, M., Zhang, F., Upton, J., Shine, P., & Shalloo, L. (2018). A review of milk production forecasting models. *Dairy Farming: Operations Management, Animal Welfare and Milk Production; Global Agriculture, Nova Science Publishers: New York, NY, USA*, 14-61.

Nanda, E. D., Salman, L. B., Indrijani, H., Tasripin, D. S., & Anang, A. (2019, October). Comparison of Five Different Lactation Curve Models to Estimate

Milk Yield of Friesian Holstein Cows at BBPTU HPT Baturraden. In *IOP Conference Series: Earth and Environmental Science* (Vol. 334, No. 1, p. 012004). IOP Publishing.

Narayan, D. (2011). *Productive and reproductive performance of Murrah buffalo breed on organized farm*. (M.Sc. Thesis. Marathwada Krishi Vidyapeeth, Parbhani, India).

Nelder, J. A. (1966). Inverse polynomials, a useful group of multi-factor response functions. *Biometrics, 22*(1), 128–141.

Nicholas, F. W., & Smith, C. (1983). Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Animal Science*, *36*(3), 341-353.

Nigm, A., Attallah, S., Abou-Bakr, S., & Sadek, R. R. (2003). A study on applying test day model for genetic evaluation of milk yield of Holstein cattle in Egypt. *Egyptian Journal of Animal Production*, *40*(2), 89-98.

Njubi, D. M., Wakhungu, J. W., & Badamana, M. S. (2010). Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Tropical Animal Health and Production*, *42*, 639-644.

Olori, V. E., Hill, W. G., McGuirk, B. J., & Brotherstone, S. (1999). Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. *Livestock Production Science*, *61*(1), 53-63.

Pander, B. L., Hill, W. G., & Thompson, R. (1992). Genetic parameters of test day records of British Holstein-Friesian heifers. *Animal Science*, *55*(1), 11-21.

Parekh, H. K. B., Singh, V. K., & Rao, K. C. (1994). Comparative study of different sire evaluation procedures under two models using crossbred progeny. *The Indian Journal of Animal Sciences*, *64*(10). 1070-1074.

Patil, C. S. (2011). *Genetic evaluation of fertility in Murrah buffaloes*. (M.V.Sc. Thesis, NDRI (Deemed University), Karnal, India).

Patil, C. S., Chakravarty, A. K., Kumar, V., Dongre, V. B., & Kumar, P. (2012). Non-genetic factors affecting first lactation reproductive traits in Murrah buffaloes. *Indian Journal of Animal Research*, *46*(2), 205-207.

Pawar, H. N., Kumar, G. R., & Narang, R. (2012). Effect of year, season and parity on milk production traits in Murrah buffaloes. *Journal of Buffalo Science*, *1*(1), 122-125.

Penchev, P., Boichev, M., Ilieva, Y., & Peeva, T. Z. (2011). Effect of different factors on lactation curve in buffalo cows. *Slovak Journal of Animal Science*, *44*(3), 103-110.

Pendor V. (2015). *Studies on reproduction and production characteristics of buffaloes at organised farm.* (M.Sc. Thesis. Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani, India).

Pereira, F. C., & Borysov, S. S. (2019). Machine learning fundamentals. In *Mobility patterns, big data and transport analytics* (pp. 9-29). Elsevier.

Pollott, G. E., & Gootwine, E. (2000). Appropriate mathematical models for describing the complete lactation of dairy sheep. *Animal Science*, *71*(2), 197-207.

Pool, M. H., & Meuwissen, T. H. E. (1999). Prediction of daily milk yields from a limited number of test days using test day models. *Journal of Dairy Science*, *82*(7), 1555-1564.

Prajapati, K. B., Tajane, K. R., Patel, J. P., & Radadia, N. S. (1992). Lactation curve in Kankrej cattle. *Gujarat Agriculture University Research Journal*, *18*(1), 90-95.

Ptak, E., & Schaeffer, L. R. (1993). Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science*, *34*(1-2), 23-34.

Pundir R K, Malik R P S and Prakash B. 2004. Comparison of different sire evaluation methods in Sahiwal cattle. *Indian Journal of Animal Sciences* 74(2): 229–31.

Pundir, R. K., & Raheja, K. L. (1994). Relationship between sires' estimated breeding value for first lactation and lifetime traits in Sahiwal and Hariana cattle. *The Indian Journal of Animal Sciences*, *64*(11), 1219-1225.

Puri, T. R., & Sharma, K. N. S. (1965). Prediction of lifetime production on basis of first lactation yield and age at first calving for selection of dairy cattle. *Journal of Dairy Science*, *48*(4), 462-467.

Quinn, N., Killen, L., & Buckley, F. (2005). Empirical algebraic modelling of lactation curves using Irish data. *Irish Journal of Agricultural and Food Research*, *44*, 1-13.

R Core Team, R. (2013). R: A language and environment for statistical computing.

Radwan, H. A. A., Abo Elfadl, E. A., & Fardos, A. (2015). Estimates of population parameters for some economic traits in Holstein Friesian cows by using statistical program. *Global Veterinaria*, *14*(1), 129-135.

Raheja, K. L. (1992). Comparison of progeny testing of Sahiwal sires by the different methods of sire evaluation. *Indian Journal of Dairy Science*, *45*(2), 64-69.

Raja, T. V. (2010). *Part lactation records for Sahiwal sire evaluation*. (Ph.D. Dissertation, NDRI, (Deemed University), Karnal, India).

Ramos, A. D. A., Malhado, C. H. M., Carneiro, P. L. S., Gonçalves, H. C., & Azevedo, D. M. M. R. (2006). Phenotypic and genetic characterization of the milk yield and calving interval in buffalo of the Murrah breed. *Pesquisa Agropecuária Brasileira*, *41*, 1261-1267.

Rana J. (2008) *Genetic evaluation of Murrah buffalo sires based on part lactation and test day milk yield records*. (Ph.D. Dissertation, Choudhary Charan Singh University, Meerut, India).

Rao, A. K., & Kumari, D. G. (2021). Performance traits of Murrah buffaloes in an organised farm of West Godavari district of Andhra Pradesh. *Journal of Entomology and Zoology Studies*, *9*(1), 256-258.

Rao, A.V., Rao, G.N. and Kumar, M. (1984). First lactation perfor- -mance of zebu, zebu × zebu and European × zebu cows. *Indian Journal of Animal Science.* 54(10): 980 - 982.

Rashia Banu, N. (2010) *Genetic evaluation of the lactation curve in Karan Fries cattle.* (Ph.D. Dissertation, NDRI, Karnal, India).

Reents, R., & Dopp, L. (1996). Genetic evaluation for dairy production traits with a test day model for multiple lactations. *Interbull Bulletin*, *14*, 113-113.

Rekaya, R., Weigel, K. A., & Gianola, D. (2001). Application of a structural model for genetic covariances in international dairy sire evaluations. *Journal of Dairy Science*, *84*(6), 1525-1530.

Robertson, A., & Rendel, J. (1954). The performance of heifers got by artificial insemination. *The Journal of Agricultural Science*, *44*(2), 184-192.

Rodrigues, A. E., Marques, J. R. F., Araújo, C. V., Camargo Júnior, R. N. C., & Dias, L. N. S. (2010). Estimation of genetic parameters of dairy buffaloes productive characteristics Eastern Amazon. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia, 62*(3), 712-717

Rubinfeld, D. L. (1998). *Econometric models and economic forecasts*. Irwin/McGraw-Hill.

Sahana, G. (1993). *Association between productive and reproductive traits in Murrah buffaloes*. (M.Sc. Thesis, NDRI (Deemed University), Karnal, India).

Sahoo, S. K., Avtar Singh, A. S., Shivahre, P. R., Manvendra Singh, M. S., Soumya Dash, S. D., & Dash, S. K. (2014). Prediction of fortnightly test-day milk yields using four different lactation curve models in Indian Murrah buffalo. *Advances in Animal and Veterinary Sciences 2(11), 647- 651*

Sahoo, S. K., Singh, A., Gupta, A. K., Chakravarty, A. K., Ambhore, G. S., & Dash, S. K. (2015). Comparative evaluation of different lactation curve functions for prediction of bi-monthly test day milk yields in Murrah buffaloes. *Animal Science*, *9*, 89-94.

Sahoo, S. K., Singh, A., Gupta, A. K., Chakravarty, A. K., Ambhore, G. S., & Singh, M. (2016). Comparison of four different lactation curve models for prediction of weekly test day milk yields in Murrah buffalos. *The Indian Journal of Animal Sciences*, *86*(1), 101-103.

Sajjad Khan, M., Hassan, F. U., Rehman, M. S. U., Hyder, A. U., & Bajwa, I. R. (2007). Genetic control of milk yield from lactations of different duration in Nili-Ravi buffaloes. *Archives Animal Breeding*, *50*(3), 227-239.

Salau, A. O., & Jain, S. (2021). Adaptive diagnostic machine learning technique for classification of cell decisions for AKT protein. *Informatics in Medicine Unlocked*, *23*, 100511.

Salehi, F., Lacroix, R., & Wade, K. M. (1998). Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Computers and Electronics in Agriculture*, *20*(3), 199-213.

Salem, M. A., Esmail, H. M., Sadek, R. R., & Nigm, A. A. (2006). Phenotypic and genetic parameters of milk production and reproductive performance of Holstein cattle under the intensive production system in Egypt. *Egyptian Journal of Animal Production*, *43*(1), 1-10.

Salem, M. M. I., & Hammoud, M. H. (2016). Estimates of heritability, repeatability and breeding value of some performance traits of Holstein cows in Egypt using repeatability animal model. *Egyptian Journal of Animal Production*, *53*(3), 147-152.

Santos, D. J. A., Peixoto, M. G. C. D., Borquis, R. R. A., Verneque, R. S., Panetto, J. C. C., & Tonhati, H. (2013). Genetic parameters for test-day milk yield, 305-day milk yield, and lactation length in Guzerat cows. *Livestock Science*, *152*(2-3), 114-119.

Sanzogni, L., & Kerr, D. (2001). Milk production estimates using feed forward artificial neural networks. *Computers and Electronics in Agriculture*, *32*, 21-30.

Sarkar, U., Gupta, A. K., Mohanty, T. K., Raina, V. S., & Prasad, S. (2006). Estimates of genetic parameters for lactational milk constituents and yield traits in Murrah buffaloes. *Journal of Dairying, Foods and Home Sciences*, *25*(2), 149-151.

Saxena, S. K., & Tomar, N. S. (1988). Genetic studies of milk production of Murrah buffaloes in first four lactations. *Indian Veterinary Journal. 65*, 925-926.

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science*, *86*(1-3), 35-45.

Schaeffer, L. R., & Jamrozik, J. (1996). Multiple-trait prediction of lactation yields for dairy cows. *Journal of Dairy Science*, *79*(11), 2044-2055.

Schaeffer, L. R., Jamrozik, J., Kistemaker, G. J., & Doormaal, J. V. (2000). Experience with a test-day model. *Journal of Dairy Science, 83*(5), 1135-44.

Seyedsharifi, R., Eskandari-Nasab, M. P., & Sobhani, A. (2008). Estimation of genetic parameters and breeding values for test-day and 305-days milk yields in some Iranian Holstein herd. *Journal of Animal and Veterinary Advances 7*(11), 1422-1425.

Seyed-Sharifi, R., Seyedsharifi, R., & Kheir, A. R. F. (2009). Selection for lactation curve and determination of the lactation curve function in Iranian Holstein cows by empirical mathematical models. *Journal of Food, Agriculture & Environment*, *7*(2), 453-455.

Shahinfar, S., Page, D., Guenther, J., Cabrera, V., Fricke, P., & Weigel, K. (2014). Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *Journal of dairy science*, *97*(2), 731-742.

Sharma, A. K., Jain, D. K., Chakravarty, A. K., Malhotra, R., & Ruhil, A. P. (2013). Predicting economic traits in Murrah buffaloes with connectionist models. *Journal of the Indian Society of Agricultural Statistics*, *67*(1), 1-11.

Sharma, A. K., Sharma, R. K., & Kasana, H. S. (2007). Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing*, *7*(3), 1112-1120.

Sharma, A. K., Sharma, R., & Kasana, H. (2006). Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. *Neural Computing & Applications*, *15*, 359–365.

Sharma, H., Gajbhiye, P. U., Ramani, A. L., Ahlawat, A. R., & Dongre, V. B. (2017). Effect of non-genetic factors on test day milk yield and first lactation 305 day milk yield in Jaffarabadi buffaloes. *Journal of Animal Research*, *7*(2), 325-331.

Sherchand, L., McNew, R. W., Kellogg, D. W., & Johnson, Z. B. (1995). Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *Journal of Dairy Science*, *78*(11), 2507-2513.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE transactions on neural networks*, *11*(5), 1188-1193.

Sigdel, A., Bhattarai, N., Kolachhapati, M. R., & Paudyal, S. (2015). Estimation of genetic parameters for productive traits of Murrah buffaloes in Kaski, Nepal. *International Journal of Research*, *2*(5), 1-11.

Sikka, L. C. (1950). 424. A study of lactation as affected by heredity and environment. *Journal of Dairy Research*, *17*(3), 231-252.

Silvestre, A., Petim-Batista, F., & Colaço, J. (2006). The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*, *89*, 1813-1821.

Singh, A., & Rana, J. S. (2008). Prediction of 305-day milk yield based on test-day values in Murrah buffaloes. *Indian Journal of Animal Sciences*, *78*(10), 1131-1133

Singh, B.P., Kumar, V. and Chauhan, V.P.S. 1992. Comparison of different methods of sire evaluation. *Indian Journal of Animal Sciences 62(7):* 749-753.

Singh, C. V., & Yadav, M. C. (1987). Effect of genetic and non-genetic factors on lifetime production traits in Indian buffaloes. *Indian Journal of Animal Sciences*, *57*(6), 555-558.

Singh, C. V., & Yadav, M. C. (1987). Nongenetic factors affecting daily milk yield in Murrah buffaloes. *Indian Journal of Animal Sciences,* *57*(1), 56-58.

Singh, M. (2014). *Genetic analysis of test-day milk yields in Murrah buffaloes.* (M.V.Sc. Thesis, NDRI (Deemed University), Karnal, India).

Singh, M., Singh, A., Gupta, A. K., Dash, S. K., Gupta, A., Sahoo, S. K., & Shivahre, P. R. (2015). Comparative evaluation of different lactation curve models in prediction of monthly test-day milk yields in Murrah buffaloes. *Journal of Animal Research*, *5*(1), 189-193.

Singh, M.K., Sharma, R.C. and Gupta, A.K. (1995). Genetic para- -meters of first lactation traits in Karan Swiss cattle. *Indian Journal of Dairy Science*. 48: 82–84.

Singh, T. P., Singh, R., Singh, G., Das, K. S., & Deb, S. M. (2011).Performance of production traits in Nili-Ravi buffaloes. *Indian Journal of Animal Sciences,* *81*(12), 1231-1238.

Sinha, R. (2019). Genetic polymorphism in VDR, PTPRR and HMGA2 gene- their association with udder type traits and clinical mastitis in Karan Fries and Sahiwal cows. Ph.D Thesis. ICAR-National Dairy Research Institute, Karnal, India.

Spearman, C. (1987). The proof and measurement of association between two things. *The American Journal of Psychology*, *100*(3/4), 441-471.

Stanton, T. L., Jones, L. R., Everett, R. W., & Kachman, S. D. (1992). Estimating milk, fat, and protein lactation curves with a test day model. *Journal of Dairy Science*, *75*(6), 1691-1700.

Strabel, T., Ptak, E., Szyda, J., & Jamrozik, J. (2004). Multiple-lactation random regression test-day model for Polish Black and White cattle. *Interbull Bulletin*, *32*, 133-133.

Strabel, T., Szyda, J., Ptak, E., & Jamrozik, J. (2005). Comparison of random regression test-day models for Polish Black and White cattle. *Journal of Dairy Science*, *88*(10), 3688-3699.

Sudhakar, K., Panneerselvam, S., Thiruvenkadan, A. K., Abraham, J., & Vinodkumar, G. (2013). Factors effecting milk composition of crossbred dairy cattle in Southern India. *International Journal of Food, Agriculture and Veterinary Sciences*, *3*(1), 229-233.

Suhail, S. M., Qureshi, M. S., Sarzamin, K., & Durrani, F. R. (2009). Inheritance of economic traits of dairy buffaloes in Pakistan. *Sarhad Journal of Agriculture*, *25*(1), 87-93.

Sundaresan, D., Eldridge, F. E., & Atkeson, F. W. (1954). Age at first calving used with milk yield during first lactation to predict lifetime production of Indian cattle. *Journal of Dairy Science*, *37*(11), 1273-1282.

Suresh, R., Bidarkar, D. K., Gupta, B. R., RAO, B. S., & Sudhakar, K. (2004). Production and reproduction performance of Murrah buffaloes. *The Indian Journal of Animal Sciences*, *74*(3), 315-318.

Suresh, R., Bidarkar, D. K., Gupta, B. R., RAO, B. S., & Sudhakar, K. (2004). Production and reproduction performance of Murrah buffaloes. *The Indian Journal of Animal Sciences*, *74*(3), 854-857.

Swalve, H. H. (1995). The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *Journal of Dairy Science*, *78*(4), 929-938.

Swalve, H. H. (2000). Theoretical basis and computational methods for different test-day genetic evaluation methods. *Journal of Dairy Science*, *83*(5), 1115-1124.

Syed, M., Zafarullah Khan, Z. K., & Shah, M. A. P. (1996). Production performance, persistency of lactation and repeatability of some quantitative traits in Nili-Ravi buffaloes. *Sarhad Journal of Agriculture, 12*(1), 1-9.

Tailor, S. P., & Singh, B. (2011). Genetic evaluation of sires using test day yield. *Indian Journal of Animal Sciences, 81*, 882-885.

Tailor, S. P., & Singh, S. (2014). Observed and predicted first lactation milk yield in Surti buffaloes. *The Indian Journal of Animal Sciences, 84*(7), 775-778.

Tailor, S. P., Banerjee, A. K., Pathodiya, O. P., & Bachchu Singh, B. S. (1998). Genetic studies of traits up to peak yield in Surti buffaloes. *Indian Journal of Dairy Science*, *51*(1), 36-39.

Tajane, K. R., & Rai, A. V. (1990). Efficiency of sire evaluation methods to improve milk yield of Sahiwal × Holstein- Friesian cattle. *Indian Journal of Animal Sciences*, *60*(1): 183–91.

Takma, Ç., Atıl, H., & Aksakal, V. (2012). Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. *Journal of the Faculty of Veterinary Medicine, Kafkas University, 18*(6), 941-944.

Talokar, A. J., Kumar, H., Mehrotra, A., Kaisa, K., Saravanan, K. A., Panigrahi, M., & Bhushan, B. (2023). Recent advances in sire evaluation methods: A review. *Indian Journal of Animal Research*, *57*(4), 395-401.

Tamboli, P., Bharadwaj, A., Chaurasiya, A., Bangar, Y. C., & Jerome, A. (2021). Genetic parameters for first lactation and lifetime traits of nili-ravi buffaloes. *Frontiers in Veterinary Science*, *8*, 557468.

Tatachar, A. V. (2021). Comparative assessment of regression models based on model evaluation metrics. *International Research Journal of Engineering and Technology (IRJET)*, *8*(09), 2395-0056.

Thiruvenkadan, A. K. (2011). Performances of Murrah buffaloes at coastal region of Tamil Nadu, India. *Indian Journal of Animal Sciences*, *81*(10), 1080-1083.

Thiruvenkadan, A. K., Panneerselvam, S., Rajendran, R., & Murali, N. (2010). Analysis on the productive and reproductive traits of Murrah buffalo cows maintained in the coastal region of India. *Applied Animal Husbandry and Rural Development*, *3*(1), 1-5.

Tonhati, H., Cerón-Muñoz, M. F., Oliveira, J. A. D., El Faro, L., Lima, A. L. F., & Albuquerque, L. G. D. (2008). Test-day milk yield as a selection criterion for dairy buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae). *Genetics and Molecular Biology*, *31*, 674-679.

Torres, M., Herv́as, C., & Amador, F. (2005). Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms. *Computers & Operations Research, 32*, 2653-2670.

Torshizi, M. E., Aslamenejad, A. A., Nassiri, M. R., & Farhangfar, H. (2011). Comparison and evaluation of mathematical lactation curve functions of Iranian primiparous Holsteins. *South African Journal of Animal Science*, *41*(2), 104-115.

Van der Heide, E. M. M., Veerkamp, R. F., Van Pelt, M. L., Kamphuis, C., Athanasiadis, I., & Ducro, B. J. (2019). Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of Dairy Science*, *102*(10), 9409-9421.

Vasconcelos, J., Martins, A., Petim-Batista, M. F., Colaço, J., Blake, R. W., & Carvalheira, J. (2004). Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *Journal of Dairy Science*, *87*(8), 2591-2598.

Visscher, P. M., & Goddard, M. E. (1995). Genetic parameters for milk yield, survival, workability, and type traits for Australian dairy cattle. *Journal of Dairy Science*, *78*(1), 205-220.

Wade, K. M., & Lacroix, R. (1994). The role of artificial neural networks in animal breeding. In *Proc. 5th WCGALP* (Vol. 22, pp. 31-34).

Wakchaure, R. S., Amit, K., Sachdeva, G. K., & Gandhi, R. S. (2008). Time series analysis of first lactation traits in Murrah buffaloes. *Indian Journal of Dairy Science*, *61*(5), 374-376.

Wiggans, G. R., & Goddard, M. E. (1997). A computationally feasible test day model for genetic evaluation of yield traits in the United States. *Journal of Dairy Science*, *80*(8), 1795-1800.

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*, *16*(4), 335-348.

Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature*, *216*(5111), 164-165.

Xu, W., van Knegsel, A. T., Vervoort, J. J., Bruckmaier, R. M., van Hoeij, R. J., Kemp, B., & Saccenti, E. (2019). Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms. *Journal of Dairy Science*, *102*(11), 10186-10201.

Yadav, M. P., Katpatal, B. G., & Kaushik, S. N. (1977). Components of inverse polynomial function of lactation curve, and factors affecting them in Hariana and its Friesian crosses. *Indian Journal of Animal Sciences*. 47(12),777-781.

Yadav, S. B., & Sharma, J. S. (1985). Functions for lactation curve in crossbred dairy cattle. *Indian Journal of Animal Sciences*. 55(1), 42-47.

Yoon, J. T., Lee, J. H., Kim, C. K., Chung, Y. C., & Kim, C. H. (2004). Effects of milk production, season, parity and lactation period on variations of milk urea nitrogen concentration and milk components of Holstein dairy cows. *Asian-Australasian Journal of Animal Sciences*, *17*(4), 479-484.