Archives of Computational Methods in Engineering (2023) 30:3031–3058
https://doi.org/10.1007/s11831-023-09892-2

REVIEW ARTICLE

Diagnosis and Detection of Congenital Diseases in New-Borns or Fetuses Using Artificial Intelligence Techniques: A Systematic Review

Komalpreet Kaur[1] · **Charanjit Singh**[1] · **Yogesh Kumar**[2]

Received: 30 September 2022 / Accepted: 29 January 2023 / Published online: 11 February 2023

Abstract
Artificial intelligence, including machine learning and deep learning, play an essential role in the medical industry for predicting various diseases. One such disease or disorder is a congenital disease that affects the newborn infant or unborn foetus by different viruses carried by the mother and passed on to the baby either during the time of pregnancy or delivery. Based on PRISMA guidelines, an extensive survey has been done to predict congenital diseases, including neonatal and postnatal. We have considered 115 articles related to the prediction of congenital diseases such as Zika virus, congenital heart disease, chromosome anomalies, sepsis, hypertension, cytomegalovirus, and many more using artificial intelligence published from 2008 to 2022 on different databases, journals, and conferences. In addition, the review also depicts the current work done by several researchers in the field of congenital disease prediction, along with their datasets and limitations. For complete work, we have designed four investigations and, in the end, explored solutions for the same. From the survey, it has been found that irrespective of various approaches used in the reported work, they can achieve predicted outcomes, but still, several problems need to be resolved. Thus, there is a need for more extensive research to deal with the challenges in the area of predicting various congenital diseases in the early stages.

1 Introduction

Congenital infections, disorders, or disease occurs when pregnant women get infected with an organism that further enters into their placenta and fetus through their bloodstream. Such disorders may affect both newborn infants and unborn fetuses. Mothers may get affected by it during the pregnancy to the time they are delivering the baby [1]. The infection that is transmitted to the child from the mother during pregnancy is known as transplacentally and on another side, one that is transmitted at the time of delivery is known as peripartum. Such infections can adversely affect the development of fetal and long-term neurodevelopmental outcomes. Commonly such diseases are referred to as cytomegalovirus (CMV), Chromosome anomalies, Zika virus, Human Immunodeficiency virus (HIV), Syphilis, Herpes simplex virus (HSV), Varicella zoster virus (VZV), Parvovirus, Varicella (Chickenpox), Congenital heart disease (CHD), sepsis, lymphocytic choriomeningitis virus (LCMV), streptococci, asphyxia, multi-organ failure, premature delivery, and perinatal demise [2]. Congenital disorders or congenital disabilities can be functional, structural, metabolic, and behavioral disorders that occur during intrauterine life and can be identified prenatally, at birth, neonatal, or later in infancy postnatal [3].

Every organ of the body can get affected because of congenital diseases [4–8]. In prevention and genetic counseling, understanding the etiology of such diseases is essential for their eradication. The etiology of birth defects is not clear but is considered multifactorial. These factors can be genetic, which counts for 10–30%, multifactorial inheritance for 20–35%, environmental factors 5–10%, and around 30–40% are unknown [9]. In medical history, the infectious

✉ Komalpreet Kaur
Chadhakomal18@gmail.com

Charanjit Singh
Channisingh@yahoo.com

Yogesh Kumar
Yogesh.arora10744@gmail.com;
Yogesh.Kumar@sot.pdpu.ac.in

[1] Department of Electronics and Communication Punjabi University, Patiala, India

[2] Department of Computer Science and Engineering, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

agents are considered the most critical environmental factors [10] that include maternal factors of age, illnesses during pregnancy, medication use, lifestyle, antenatal care, and non-use of peri-conceptual folic acid [7, 9, 11, 12]. In the aetiology of congenital diseases, other vital factors are previous miscarriages, parental consanguinity and stillbirths, and inheritable congenital diseases [7, 11, 12]. Pregnant women must avoid contact with people who are not well, especially the women who are seronegative or have no previous exposure to any disease. They also need to avoid eating raw or undercooked food such as pork, beef, and lamb that contain *T. gondii*, Campylobacter fetus, Listeria monocytogenes, and *Salmonella* spp. Pregnant women must also avoid sexual intercourse with their partner if they have HIV or genital herpes. Breastfeeding to the newborn in HIV should be avoided for symptomatic mothers and one having low CD4 T cell counts to lower down the chances of child gets infected following birth. Even the pregnant women who are culture positive for Group B Streptococcus during delivery and labor need to be treated. News eyes with erythromycin need to be treated to avoid ophthalmic neonatorium.

If the women trying to become pregnant are seronegative, then immunizations of VZV, Hepatitis, and Rubella need to be given to them. Before 3–6 months of conception, live viral vaccines should be given to women. Still, seronegative women who are already pregnant do not get cured by giving any live viral vaccines to them [13]. Artificial Intelligence (AI), including Machine learning (ML) and Deep learning (DL), has become a necessary part of the medical industry for predicting various diseases from the database related to the medical field. Many firms are employing such approaches for the early prediction of illness and diagnosing it medically in a better way [9]. One such disease or disorder as shown in Fig. 1 is a congenital disease that can be perinatal and postnatal and can affects the newborn infant or unborn fetus by different viruses carried out by the mother and passed on to the baby either during the time of pregnancy or delivery. Extensive research shows very high risks of diagnosing it wrong due to poor investigation and irregular follow-up during pregnancy [5]. The introduction of AI in prediction modeling, early detection, monitoring, and diagnosis of congenital diseases can help save the lives of newborns and premature babies. The idea behind using AI in various fields is its simplicity, and there is no need to use any mathematical function to model the scenario. ML and DL have also played an essential role in different areas of medical science. Even an improved result has been seen in the prediction of congenital diseases. So, the motive behind this study is to review various ML and DL techniques that has been already employed in the prediction of various congenital diseases.

2 Contribution and Organization of Paper

We extensively surveyed various ML and DL methods to predict congenital diseases, including neonatal and postnatal [8, 14, 15]. This paper presents a comprehensive analysis of other research on predicting such disorders or diseases using AI based techniques and the dataset used for it. Most of the existing approaches proposed by different researchers based on a deep learning framework can give appreciable outcomes. This paper provides a description of different types of congenital diseases along with illness systems and their causes, further giving brief detail about various congenital diseases with ML and DL models used to predict the same.

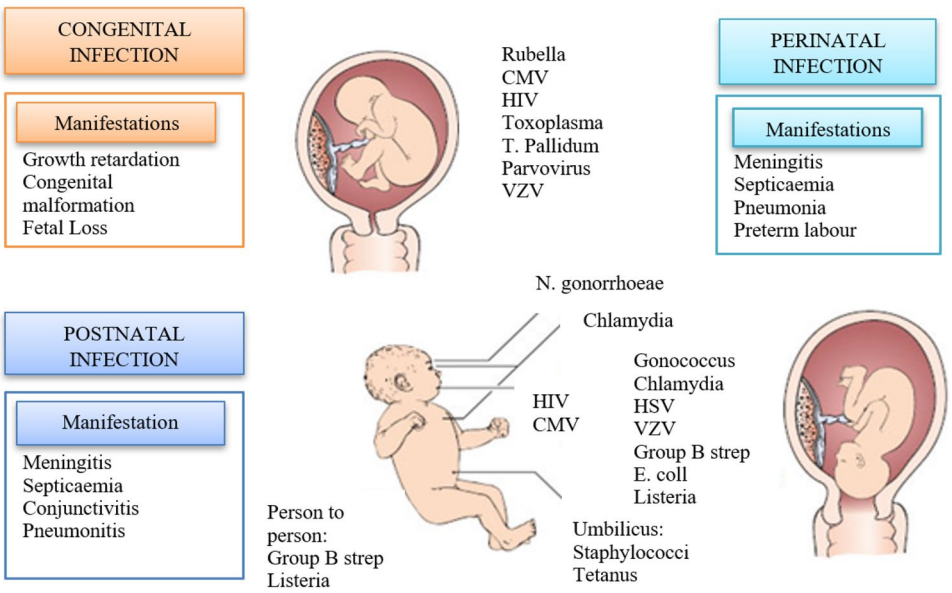

Fig. 1 Congenital infection [4]

The following section highlights the limitations of existing models based on a comparative study done on existing work and the metrics used for analysis.

3 Research Methodology

In this study, we have conducted a systematic review based on Preferred reporting items for systematic Reviews and Meta-Analyses (PRISMA) guidelines. We have considered articles on congenital diseases published from 2008 to 2022 in different databases, Journals, and Conferences [7, 8, 11, 16–18]. We have studied a total of 180 papers. Some of the latest articles' access was unavailable, so we shortlisted them, then the review and duplicate articles were skipped out from the final list of the papers used in this paper.

We efficiently searched for a selection of research papers on Google scholar, IEEE Xplore, Research gate, and science direct. Some of the papers were available openly, and few needed to purchase from sites to get full access to the paper. The papers and articles were selected using the query ((Machine learning) or (Deep learning) or (Congenital disease) or (prediction) or (name of different congenital diseases)) [19–21]. Further, articles were selected using exclusion and inclusion standards discussed in Sect. 2.1 and Fig. 2, representing the PRISMA flowchart depicting the detailed screening of the collected, accepted, and rejected papers. The papers published between 2008 and 2022 have been included in this study. A total of 180 studies were selected, and we left with only 150 papers after removing the duplicate ones. Finally, the last 115 papers were established and analysed by focusing on diseases other than congenital and papers in language other than English were excluded from this review article. After that complete shortlisted papers were evaluated and removed the articles in which other than ML or DL technique/ only genetic algorithms were used. In the end, total 115 selected papers and articles were analysed thoroughly with complete details.

3.1 Quality Assessment

While including papers in this review work, some quality evaluation restrictions have been considered. The inclusion and exclusion standards were considered for assessing the significance of the study. Most of the papers including in the paper, contains prediction based on ML and DL for different congenital diseases [20]. Every paper considers for the survey consists of empirical research with experimental outcomes. Initially, we studied various papers related to congenital diseases and then shortlisted some review papers

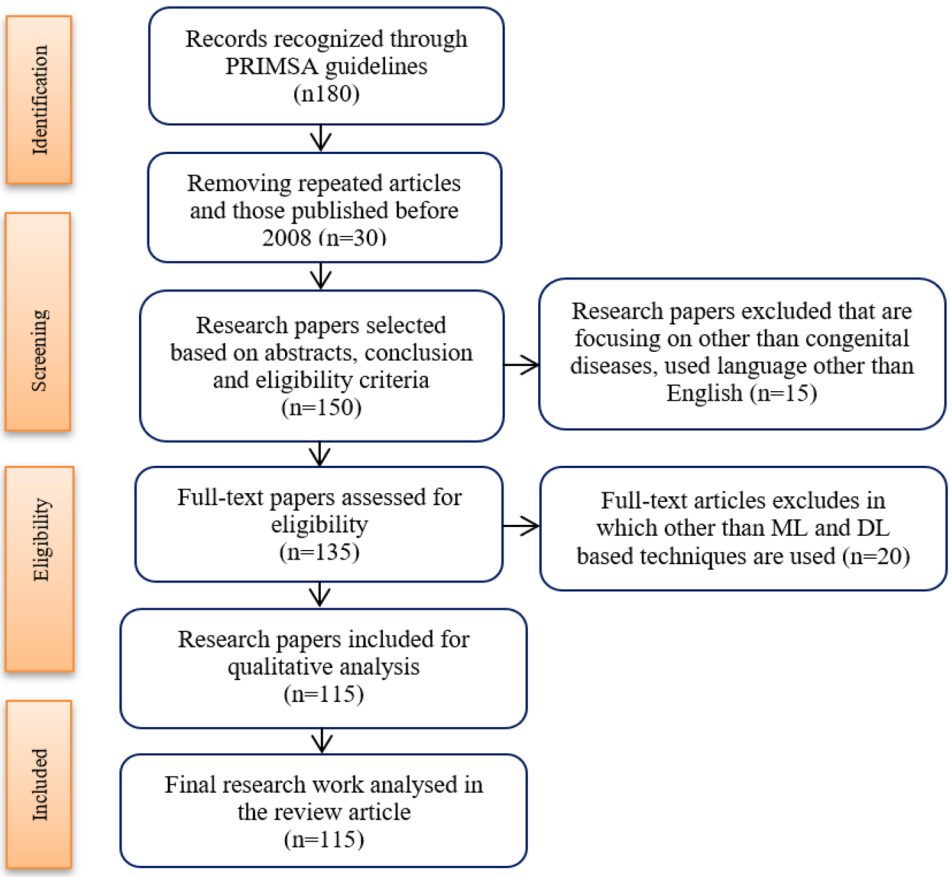

Fig. 2 PRISMA Flowchart

to get the introduction and other general details about multiple diseases. Further, we have shorted the best papers in which researchers have used ML or DL for the prediction of various diseases comes under congenital conditions different categories. For this, we have studied the abstract and conclusions of different papers and aggregated all the research retrieved from them. The inclusion and exclusion method used for this paper is shown in Fig. 3.

3.2 Investigations

In this review work, we have done the following investigations:

Investigation 1: Years-wise publication of different congenital diseases.
Investigation 2: Which ML and DL techniques have been used for the prediction of congenital diseases?
Investigation 3: What are the limitations of existing congenital disease prediction models using ML and DL techniques?
Investigation 4: What metrics are used for evaluating ML and DL models?

4 Framework for Predicting Congenital Disease

This section details the general framework followed to predict congenital diseases using ML or DL models [11]. As shown in Fig. 4, in the first step data repository is chosen, referred to as a data archive. In this area, researchers can get information or material of their interest. The most efficient way of publishing and disseminating research data is through a research data library that can save lives, provide solutions, and increase the viewer's knowledge.

After the first step, qualitative or quantitative data can be taken from different sites, sources or users while performing experiments with relevant data [15]. Data collection based on experimental methods is a systematic approach in which data is collected and analyzed from different sources to achieve a realistic and holistic view of an interesting field.

After that, the required data will be selected based on a subset of the accessible data to work. This step may contain unnecessary data such as duplicate, noisy and missing that will be cleaned out in the same step to solving the present issue. There can be the possibility of eliminating the occurrences [15, 22]. An outcome of previous steps may be sensitive information that needs to be deleted completely. In data pre-processing, the unstructured data is transformed into a usable format as accurate world data is frequently imprecise. The data must be adequate to train models for better outcomes [9].

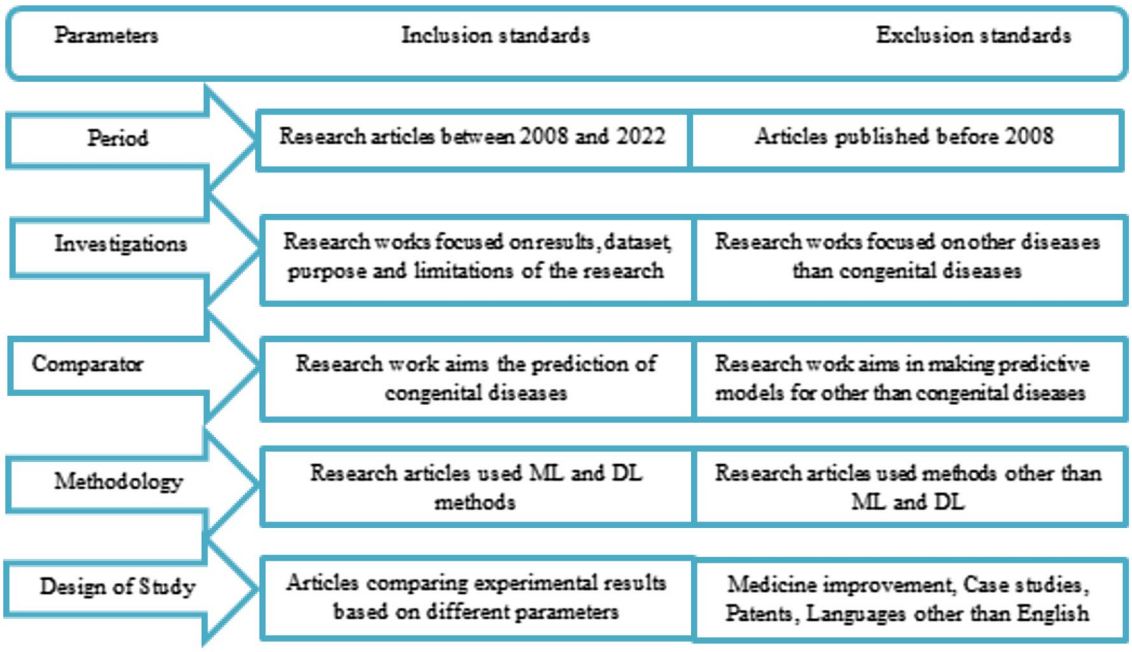

Fig. 3 Inclusion and Exclusion standards

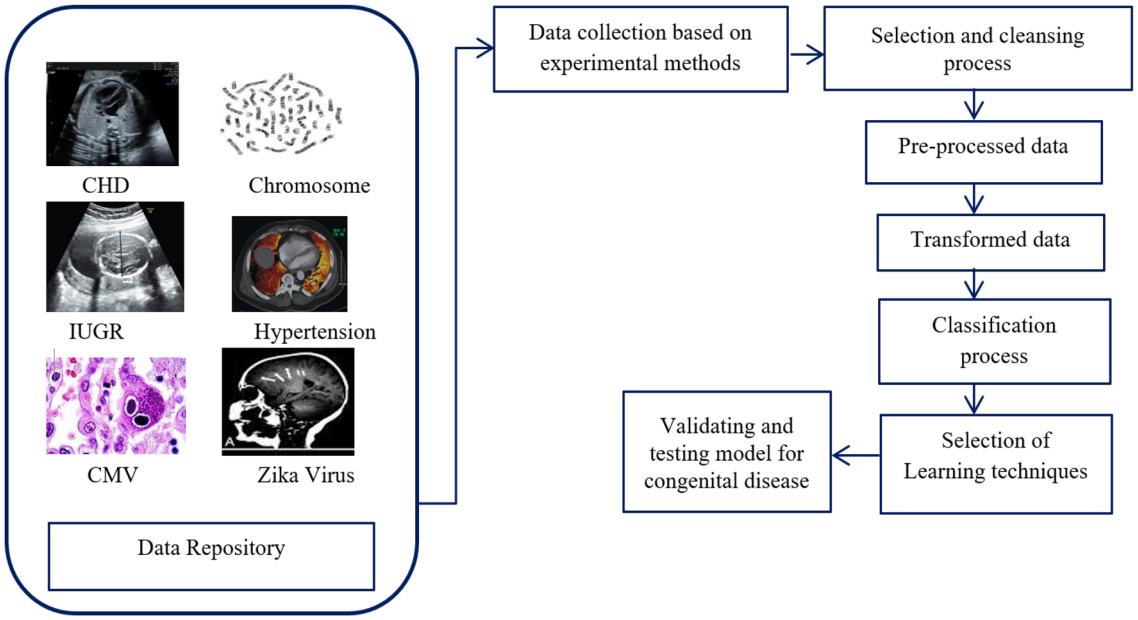

Fig. 4 Framework for predicting congenital disease [15]

Further, the classification step takes place on the transformed data that classify collective data into different categories using an appropriate classification model based on the type of dataset used for the prediction. For instance, in the case of neural networks, there is a need for extensive data, and ensemble learning-based classification methods give a good outcome in the case of unbalanced datasets. The decision tree model provides better output if data contain closely and strongly correlated characteristics. Last, the validation and testing step took place that evaluate the trained model based on testing dataset and validation results help to identify the model with best performance.

4.1 Various Congenital Disease or Disorder

There are several congenital diseases that can be neonatal, prenatal and postnatal. In the Table 1 given below, we have tried to cover most of the common and uncommon viruses that can affect a fetus and new-born baby. Also, different disease has its own effects on the fetus and new-born infant that is also tried to highlight in Table 2.

In the above Table 1, we have covered congenital diseases such as CMV, Hepatitis B, Varicella-Zoster Virus (Chickenpox), IUGR, Zika Virus, Chlamydia, Sepsis, Streptococcus, CHD with their illness to infants, causes and its etiologic agents. For more details, we have added an image available on the internet for such diseases in babies or infants. Fetus infect can lead to death and resorption, stillbirth and abortion of the fetus. There can be a premature baby live birth who may not be normal [21, 46, 47]. As shown in the Table 2, the utero infections like low birth weight, development abnormalities and congenital disease can appear in the infant. The utero acquired infects may continue even after birth and can results in abnormalities in them after some duration of time of the birth or may not be able to recognize it for months or a years.

5 Reported Work

This section covers the investigation part 2 that is to give a detail about various ML and DL techniques used for prediction of congenital diseases. We have tried to cover maximum congenital diseases namely fetal health, chromosome anomalies, Congenital heart disease (CHD), Cytomaglovirus (CMV), Zika virus, Sepsis and Preeclampsia (PE) hypertension [48–50]. We have tried to cover maximum papers of existing work done in these diseases using ML and DL models with outcome achieved by individual.

5.1 Prediction of Fetal Health

Nowadays, deaths related to pregnancy and increase in complications related has made it a big medical global issues. Internationally, there are about 830 maternal deaths on daily basis that are related to pregnancy and childbirth related complications. Mostly, this took place in developing countries [51]. So, Haghpanahi et al. have developed an algorithm maternal and fetal subspaces iterative decomposition based for automatic locating QRS complexities

Table 1 Various Congenital disease or disorder [5, 6, 9, 10, 13, 14, 16, 17, 23]

Disease	Illness to infant	Causes	Etiologic agents	Image
Cytomegalovirus (CMV) infections [19, 12]	Illnesses to infant due to CMV are broken blood vessels, hearing loss, Jaundice, Small head and large liver	It spread from person to person by coming in contact with infected urine, vaginal fluid, saliva and semen	Virus	
Chlamydia [24]	Illnesses to infant are erysipelas, rheumatic fever, strep throat, scarlet fever, glomerulonephritis and strep pneumonia	It gets transmitted sexually and can impact both males and females	Bacteria	
Hepatitis B [20, 25]	It causes advanced damage to the life throughout the life of a child with symptoms of joint pain, low-grade fever, nausea, fatigue, dark color urine, jaundice, muscle pain and loss of appetite	It took place by coming in contact with the bodily fluid and blood of an infected person	Virus	
Herpes simplex virus	This disease localized to the eye, mouth, and skin. In the worst case it can affect multiple organs, liver, eye, lung and many more	This infection is present in one in a 5 women who are pregnant	Virus	

Table 1 (continued)

Disease	Illness to infant	Causes	Etiologic agents	Image
Parvovirus B19	Illnesses to the infant contain joint pain, swelling, runny nose, headache, rash and fever	It occurs when a child comes in linked with a person who is suffering from parvovirus B19	Virus	
Rubella (German measles)	Illnesses to the infant contain delayed growth, heart defect, enlarged liver, diabetes, cataracts, low birth weight, hearing problem and mental disabilities	This virus is spread through bloodstream, bodily fluids that can be from the nose and throat	Virus	
Pediatric syphilis	Illnesses to the infant contain muscle pain, weight loss, fever, sore throat, hair loss, swollen lymph glands and headaches	It is a sexually transmitted disease in which pregnant woman suffering from syphilis spread the disease to her baby by the fetus during childbirth	Bacteria	
Varicella-Zoster Virus (Chickenpox)	It causes Eye abnormalities and Skin Scars		Virus	
Zika Virus [26–28]	Symptoms of the Zika virus are mild and it can cause pregnancy loss and other complications related to pregnancy	It is spread by the bite of a certain type of mosquito and sexual contact with the person suffering from the zika virus	Virus	

Table 1 (continued)

Disease	Illness to infant	Causes	Etiologic agents	Image
Intrauterine growth restriction (IUGR) [29–31]	Inadequate nutritional milieu Low birth weight			
Sepsis [32–36]	Symptoms are not specific but some of the common are: Jaundice, lethargy, breathing problems, low or unstable body temperature, Lower heart rate	It can be caused by fungi, bacteria, viruses and parasites	Infection	
Streptococcus	Symptoms are Fever, Tiredness, Weakness, Respiratory problems, Weight loss, heart function related problems	It can be caused by direct contact with an infected persons skin, mucus, infected lesions spreads from person to person	Bacteria	
Congenital heart disease [37–45]	Symptoms are blue colour skin, lips, nails and finger, poor blood circulation, fatigue, short and rapid breath	It is the problem in structure of infants heart, arteries, veins, wall in the heart or disruption in the normal flow of the blood through heart	Disease	

Table 2 Effect of various diseases on the fetus and newborn infant

Disease	Prematurity	In utero growth retardation and low birth weight	Developmental anomalies	Congenital disease	Persistent postnatal infection
Rubella	–	+	+	+	+
CMV	+	+	+	+	+
HSV	+	–	–	+	+
VZV	–	(+)	+	+	+
Enteroviruses	–	–	(+)	+	–
Hepatitis B	+	–	–	+	+
HIV	(+)	(+)	(+)	+	+
Erythrovirus B19 (Parvovirus B19)	–	–	–	+	–

in non-invasive fetal ECG signals taken from a set of four electrodes placed on the mothers abdomen [52]. After removing maternal components, the signals are merged using novel merging technique and detected the fetal QRS complexes from it. The PhysioNet/CinC challenge 2013 provided fetal ECG datasets were used to train and test the algorithm. The final outcome shows that the proposed algorithm was able to detect fetal peaks for different morphologies signals and strength levels in clinical practice. Further, Abbasi et al. have worked on detection of sharp wave transients of later phase of hypoxia–ischemia (HI) in the ECG to determine its potential [53]. For this, they have developed a novel computational EEG analysis method based on type-2 fuzzy logic system classifier. Multi input single output (MISO) rule of Type-2 fuzzy for the detection of sharp waves is represented in Eq. 1 taken from [53]:

$$iff\ (LMF) \leq f(W(signal)) \leq f(UMF), \tag{1}$$

Then Class is 'Sharp wave detected'.
Else Class is 'Not Sharp wave.

In above equation, LMF and UMF stands for lower and upper Type-2 input MFs respectively. After testing the proposed model, they were able to achieve the overall performance of 94% for clinical 64 Hz sampled EEG.

Miao et al. have proposed an alternative and better AI based approach for cardiotocographic diagnosis of fetal based on multiclass morphologic pattern predictions contains 10 classes of target using DL models [54]. The results obtained from it show that the proposed deep neural network model able to give accuracy of 88.02% with precision, recall and F-score of 85.01%, 84.30% and 85.08% respectively. In the same year Akbulut et al. have worked on congenital anomalies that has been seen in around 1–3% of the population that is diagnosed with double, triple and quad tests during pregnancy [55]. In this topic, Akhan has proposed a prediction system with assistive e-Health applications that can be used by both practitioners and pregnant women. In their work, they have used Boosted DT, Bayes Point Machine, Decision Forest, Decision Jungle, Locally-deep SVM, LR, NN and SVM binary classification models. These all models were trained with the 96 pregnant women clinical dataset taken from Radyo Emar radio diagnostics center in Istanbul, Turkey. Their applications able to give the status of pregnant women health and input of clinical history parameters give recommendation of physical activities that need to perform during pregnancy and inform patients about anomalies as the output [56]. The highest prediction accuracy was achieved using decision forest model of 89.5%. After that, Pisapia et al. have worked on the determining various extracted image features from MRI of fetal and integrated it using ML techniques that enable to predict the requirement of postnatal cerebrospinal fluid (CSF) diversion after birth [57]. They have considered total of 74 patients and classification results achieved were CSF diversion accuracy of 82%, Sensitivity of 80%, specificity of 84%.

5.2 Prediction of Chromosome Anomalies

For its accurate diagnosis of down syndrome, Feng et al. have utilised deep learning techniques for building accurate models for prediction and screen based on newly induced Illumina genotyping array analysis [58]. Vanderbilt University Medical Center collected clinical genotyping data based chromosome SNP maps were built by them after that they have proposed CNN architecture with two input chromosome SNP maps. With the proposed technique they were able to achieve the average accuracy of 99.3%. Further, Neocleous et al. have introduced non-invasive diagnosis procedure for reduce costs occur socially and financially in diagnosis tests of prenatal performed in an early trimester of pregnancy [59]. For this, they have proposed a method based on ANN that was trained with singleton pregnancy cases datasets namely 122,362 euploid and 967 aneuploid cases. They have collected data

from mother and the fetus aswell that is further used to train ANN through markers value in their raw form. In their study, they have firstly applied k-means unsupervised clustering algorithm to euploid cases with five prototypes. This count was selected using Elbow method then k-nearest neighbors of the respective prototype was identified for every cluster. The quantity of k proportional to the length of the respective cluster with total euploid population's length that is automatically identified and represented using below Eq. 2 taken from [59]:

$$k = targe\_population \frac{size\left(cluster_k\right)}{size(euploid)}. \quad (2)$$

The above equation was used to present the number of representative cases collected from every cluster. Through their proposed models they were able to achieve the detection rate of Trisomies 12, 18 and 21 is 80%. Later, in 2018, Somasundaram, has proposed method for chromosome classification based on deep features using CNN. For this, they have trained their model on different chromosome datasets consist of adaptively resampled image patches [60]. The results gives classification accuracy of 98.7% with 0.97 values AUC and 98.4% accuracy of abnormality detection. Visual search and identification of human being chromosomes has become necessary clinical procedure for screening and diagnosing genetic disorders and cancers. In this, karyotyping and abnormality detection on the karyograms are most common steps. Karyotyping is a standard approach used for classifying metaphase chromosomes into 24 types called karyograms. The manual genetic diagnostics is a labor intensive and time consuming task that has attracted various researchers in developing automatic computer assisted genetic diagnostics systems. While there are numerous methods for automated segmentation and classification of chromosomes, karyotyping analysis remains challenging. The flowchart used by Devaraj in his proposed work is given below in Fig. 5:

For estimating the fetal trisomy 21 risks along with the risks of other chromosomal abnormalities took place at the 1–13 gestations week of computational intelligence classification method were used by Neocleous et al. They have divided dataset into different stages and results obtained at stage 2 were 97.1% sensitivity and 99.5% of specificity [61]. Then, Sharma et al. have proposed Residual convolutional recurrent attention neural network (Res-CRANN) that uses chromosome classification band sequence properties. They have demonstrated the proposed model efficacy on a Bio image chromosome classification dataset available publically and found that their model was able to perform better than baseline models based on traditional deep CNN and ResNet-50 by around 3%. In 2019, Qin et al. has presented a novel Varifocal-Net approach for simultaneous classification of type and polarity of chromosome using deep CNN [62]. The work was divided into three stages based on one global scale and one local scale network approaches. The proposed approach was tested on dataset and found that out of 1909 karyotyping cases the proposed approach was able to achieve highest accuracy of 99.2% for both type and polarity tasks. Then again in the same year Dr. Jaganathan et al. have also worked on the detection of DS genetic disorder by initiating an accurate and non-invasive diagnostic process using ML approach with the motive of reducing cost of basis prenatal diagnosis [63]. They have used L1-norm based SVM for feature selection and results obtained from it show that the proposed LA normal SVM able to fulfil the gap between the process of feature selection and classification.

In cytogenetics lab, analysis of chromosome is essential in which abnormalities are diagnosed by cytogeneticists. So, to automate karyotyping for recognizing the common numerical abnormalities a DL was applied by Al-Kharraz, et al. on 147 non-overlapped metaphase images dataset taken

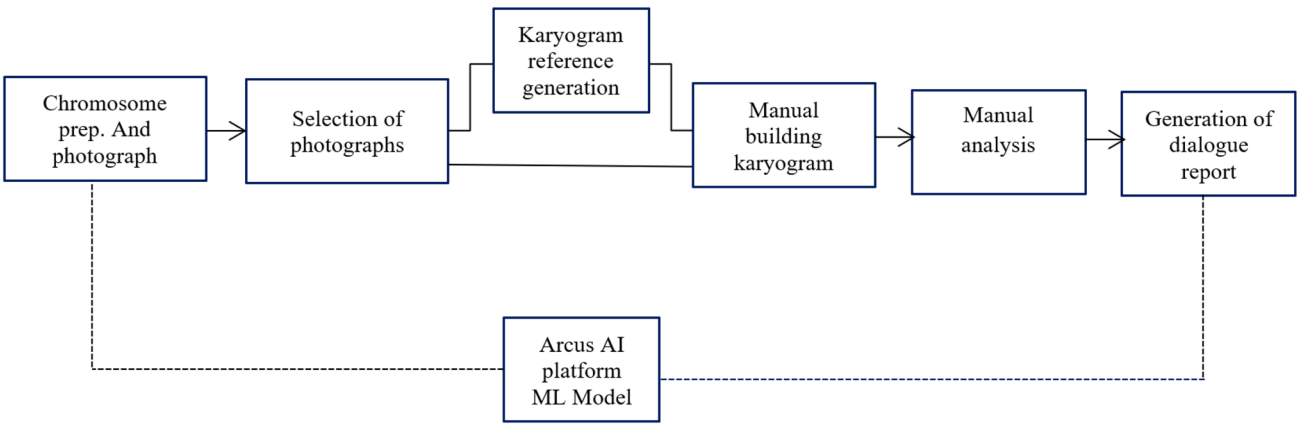

Fig. 5 Flowchart of Karyotyping and abnormality detection [60]

from King Abdulaziz university [64]. There are three stages through which metaphase images are sent, in the first stage the individual chromosome were detected using YOLOv2 CNN that give means IoU of 0.84 and accuracy of individual chromosome detection was 100%. In the second stage, VGG19 network was used that give accuracy of 95.04% and final stage of detecting abnormality gives the abnormality detection accuracy of 96.67%. Chromosome anomalies are another one of the main problem seen in fetus and infants. For diagnosing fetal chromosomal abnormalities Xu et al. have explored soft ultrasound market (USM) and non-invasive prenatal testing (NIPT) [65] using ML and data mining techniques. For analysing the ultrasonic examination data taken from 856 cases having high risk single pregnancy during early and middle stage of pregnancy they have applied NIPT. The examination results show that using proposed model one can get 98.29% accuracy, 96.72 and 98.45% of sensitivity and specificity respectively. In 2022 Nimitha et al. have worked on detection of chromosome anomaly using neural nework based VGG16 DL for transfer learning with a goal of dividing karyogram into normal and abnormal categories and give confidence level [66]. Further, abnormal is categorised into structural and numerical one. But in the proposed work, they have only worked on detection of numerical abnormalities through which they will be able to identify chromosomes which are either missing or extra in count from the original pairs of 23 one [67]. For testing the proposed model, they have used Biomedical Imaging Laboratory dataset that gives the accuracy of 98.6% and precision of 95.5%.

5.3 Prediction of Congenital Heart Diseases (CHD)

CHD is considered as one of the most common birth defects in different nation of the world. Its risk factors has been examined by various researchers but its predictive abilities have not been evaluated. So, Luo et al. have proposed and validated ML model for predicting risks bearing children of CHD before and during pregnancy [68]. The Weighted SVM, Weighted RF was trained on comprehensive population based retrospective cross sectional epidemiological birth defects survey of six countries in china during the period of 2006 to 2008. and LR was fitted to two third of the data that was able to give outcome of 0.65 True positive rate and 0.93 true negative rate. Further, Li et al. have used ANN model for prediction of risks associated with CHD in pregnant women [37]. They were able to achieve the sensitivity of 78%, specificity of 90% and Yuden Index of 68%. The existing study shows that there are more than 50% of newborns suffering from CHD that is unrecognized at their birth. So, the diagnosing of CHD incidence identification was performed by Yoon et al. using Echo and systemically review on recommendation of such group of patients [38]. They have reviewed around 630 studies with 1928 new-borns with ANCM and 719 that is 37.3% were diagnosed with Echo as having CHD. Du et al. have used child ECG from 68,969 patients at the GZFEZX to train deep NN model with two-level resident for identification of CHD patients based on it [39]. The experiment result shows that 92.30% of accuracy has been achieved using proposed model with 74.73% of sensitivity and 94.07% of specificity.

In 2020, Meda et al. has explored the best approach for non-clinical information using various ML algorithms such as Naïve Bayes, Feedforward ANN [40]. The experiment results show that ANN model was able to give the accuracy of 99.6% with 90.9% of impressive weighted accuracy on the selected dataset. Arnaout et al. have worked on ultrasound 107,823 images taken from 1326 retrospective echocardiograms and screening ultrasounds from 18 to 24 week foetuses [41]. Using it, an ensemble NN was trained for the identification of recommended cardiac views and distinguishes between normal and complex CHD. The experiment performed on it gives the accuracy of 99% with sensitivity and specificity of 95%. After John, computer aided fetal heart echocardiography examinations based on deep learning model was proposed by Nurmaini et al. They have also used instance segmentation approach to inherently segment four standard heart views the detect defects from it [42]. The used instance segmentation approach was developed using the concept of Mask-RCNN architectures having two main processes, region proposal networks as feature extraction and as multi-task learning process they have used fully connected networks in terms of classification, detection and segmentation as shown in below Fig. 6.

They have used 1149 fetal heart images for experiment purpose and predicted 24 objects containing 4 shapes of fetal heart standard views along with in each view there were 17 heart chambers objects and three cases of CHD [69]. The results obtained from it show the satisfactory performance with 98.03% of average precision for intra-patient variation and 82.42% for inter-patients variation. With the concern of increase in CHD cases, Ammarah et al. have also proposed model for early prediction based on different factors extracted from medical research [43]. Using these factors, a questionnaire was designed and applies different ML models such as KNN, DT, RF, SVM and deep learning model MLP. Among all models, MLP classifier was able to give best results with 92.9% of accuracy. Later on, Bharti et al. have also applied different ML and DL algorithms and compare the results and analysis of the UCI ML heart disease dataset that gives the accuracy of 94.2%. Truong et al. have achieved sensitivity of 85%, specificity of 88% and negative predictive value of 97% [44]. A birth cohort study at China in one of the largest cardiac centers was conducted by Qu et al. on the period of data between 2011 and 2017 [45]. They have screened all foetuses for CHDs using ultrasound

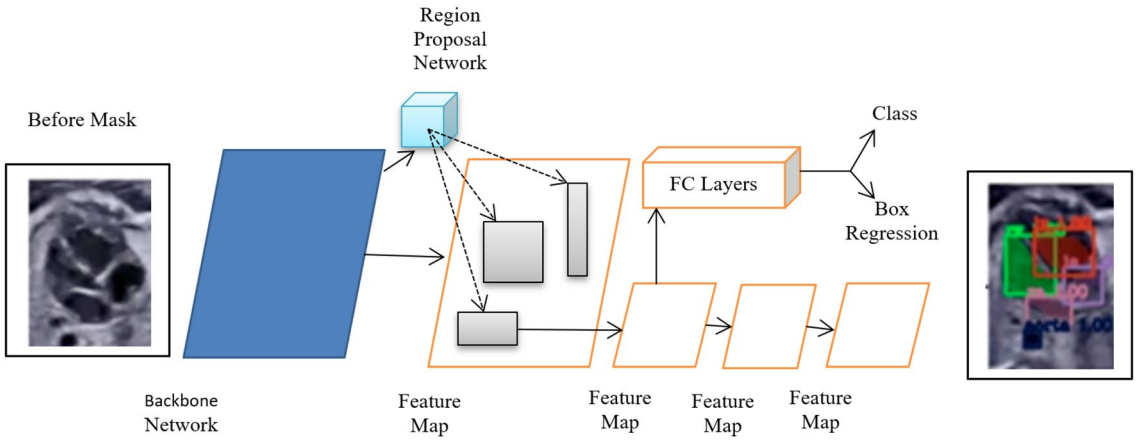

Fig. 6 Instance segmentation approach [42]

and confirmed cases by at least two paediatric cardiologist using ECG. For the prediction, they have used explainable boosting machine and able to achieve accuracy of 65% with sensitivity and specificity of 74 and 65% respectively.

5.4 Prediction of Intra-uterine Growth Retardation (IUGR)

The origin of IUGR is unknown also it is in most interest as now Fetal Origin of Adult Disease is able to recognise in better way. So, a benchmark based on existing analysis IUGR dataset was built by Buscema et al. that consists of 46 subjects described into 14 insulin like growth factor system related variables [70]. For their work, they have utilised new algorithms based on two NN namely Auto contractive Mao and Activation or competition system algorithms. In their work, they were able to identify the variables which were predictive of fetal growth. In IUGR prediction, Wosiak, et al. have proposed automated ML based model for identification of IUGR [71].

In practice plurality voting regrading classifier combination rules was used for implementation except simply majority and unanimity. In work done by Bahado-Singh et al. a pathogenesis of IUGR were interrogates and applied AI techniques to various platforms such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) based metabolomic analysis [72]. With the use of SVM, they have achieved the range of accuracy between 78 and 82% that was best compared to other algorithms.

IUGR and preeclampsia are placental dysfunction related disorders that need referral decision made in certain period of time. This creates a need of appropriate prediction models for such diseases. So, PDDs predictive model using ML was developed by Sufriyana et al. that consists of 24–37 weeks gestation features of various other characteristics of maternal [73]. For testing the proposed model they have consider pregnant women with PDDs (66/95, 69%) and a control group (29/95, 31%) dataset were utilised. Final outcome shows MCC value of 87%, specificity of 95% and sensitivity of 95% that is better compared to other recent studies they have considered for comparison purpose. In the prediction of IUGR, Pini et al. was able to achieve the accuracy, sensitivity and specificity of 93%, 93% and 84% respectively using radial basis function support vector machine (RBF-SVM) mode [74]l. In the same year Crockart et al. have developed a new ML model for accurately predicting the estimated featl weight at 34 + 0–37 + 6 weeks gestation stage is below 10th percentile using data collected at 20 + 0 to 23 + 6 weeks gestation [75]. For analysis they have used stochastic GD, KNN, LR and RF methods and was able to achieve the accuracy of 93% using first ML method. In year 2022, Teng et al. and Aslam et al. also used ML in a better way for predicting IUGR in fetus [76, 77]. The Doppler velocimetry along with second-trimester ultrasound biometry was used by Lung Yun for constructing six FGR ML models. The multicollinearity test on the data was used for feature correlation and correlation plot was constructed using gcorrplot in R function for visualising the collinearity among the predictor variables. The below given Eq. 3 was used to compute the correlation matrix on all the predictor variables [77].

$$r_{jk} = \frac{S_{jk}}{S_j S_k} = \frac{\sum_{i=1}^{n} \left(x_{ij} - x_j\right)\left(x_{ij} - x_k\right)}{\sqrt{\sum_{i=1}^{n} \left(x_{ij} - x_j\right)^2} \sqrt{\sum_{i=1}^{n} \left(x_{ij} - x_k\right)^2}}. \tag{3}$$

In the above equation, r_{jk} is a correlation between predictor j and k, S is the variance, S_{jk} is the covariance between predictor j and k, x is the predictor variable vales and x with single variable is the mean of that predictor variable [78]. The results of his study show a better outcome using SVM. On other hand, Nida has utilised SVM along with other ML algorithms namely RF, KNN and Gradient Boosting. For the

purpose of fetus classification a Recursive Feature Elimination (RFE) method was used and able to achieve best outcome using RF with 97% accuracy and 98% F1-score.

5.5 Prediction of Cytomegalovirus (CMV)

CMV is the most important agent of congenital infection in infants that consider as the main cause of not genetic sensorineural hearing loss (common congenital kind of hypoacusia). Carnimeo et al. have used neural diagnostic support tool for identification of children from clinic, virologic, anamnestic and serologic point of views [79]. The use of intelligent analyser is able to give valid contribution in the early diagnosis of this deep form of hypoacusia. The count of CMV strains in human population is getting increased due to which the biological significance of these strains has been an active area of research [80]. So, to determine the feasibility of predicting congenital CMV disease an ANN model was constructed by Arav-Boger et al. and applied it on two datasets [81]. First dataset was 54 sequences of CMV gene UL144 taken from 54 women amniotic fluids who contracted acute CMV infection during the period of their pregnancy along with US28, UL144, UL146 and UL147 4 genes 80 sequences obtained from urine, saliva or blood of infected infants. The experiment results on it show the prediction outcome of 80% for UL144 and 85% for UL146. In early diagnosis of CMV, Tanimura et al. have found that if in maternal uterine cervical secretion a CMV DNA is present then there are chances of congenital CMV infection in pregnant women who are CMV immunoglobulin M (IgM)-positive [82]. For this work, they have taken review of current knowledge of the predicting congenital CMV infection potential biomarkers. After that Rogers et al. have descried clinical experience using CMV T cell Immunity Panel (CMV-TCIP) for the first time that comprised of assay for measuring D4+ and CD8+ Tcell responses, for prediction of CMV events clinically [83]. They have nearly analysed 44 samples from 37 patients in which 31 recipients were solid organ transplant, 4 had hematologic malignancies and 2 had autoimmune disorders. With this, they were able to get the accuracy of 78% for CD4+ and 66% for CD8+ T cells. To take the benefit of digital image processing performing good in improving the quality of the images, Deepa et al. have taken the image of CMV and processed it to get improved outcome from affected images [84]. In their proposed work, they have processed input image with ROI and MAFE algorithm along with PCA reduction. For final classification, HSVM was used perform the identification of image that give better results. In the same year, Hu et al. have proposed and tested DCNN for analysing end to end fashion cytometry data [85]. In the same work, they have interpreted the model using permutation based method and results show that the proposed model was able to identify CD27–, CD94+ and CD8+ T cell population significantly associated with latent CMV infection. After a year, Lee et al. have developed a ML based classifier for differentiating between CMV and HSV esophagitis [86]. With this, Jung was able to get 100% of accuracy along with other associated parameters. Later on, Eisenberg et al. have also developed ML models integration of data of baseline patient and time dependent laboratory measurements for individual prediction of mortality and CMV reactivation after hematopoietic cell transplantation (HCT) at multiple time points per patient [87]. The outcome obtained fro test depicts that GB ML models able to give time dependent, well calibrated, risk predictions with AUC of 0.92 and 0.83 and prediction recall of 0.58 and 0.62 prediction of mortality in 21 day time period.

5.6 Prediction of Zika Virus

Zika virus is another disease that is linked to severe congenital abnormalities cause global health issues. So, for mapping the probability of Zika epidemic a backward propagation neural network (BPNN) along with RF and GBM was used by Jiang et al. [88]. Based on training dataset, 50 modelling processes were conducted to evaluate the performance of ML models that show a best outcome is obtained using BPNN that is 95% of accuracy. A year later, Mahalakshmi et al. have worked on developed an early prediction system using symptoms for detecting the virus [89]. To improve the classification accuracy the time for taking number of inputs and training it get reduced.

$$Net_j = \sum W_{ij} X_j + \theta_j. \tag{4}$$

In the above Eq. 4, w is stands for the weight of each node, x is the i to j input value [89]. Next the sigmoid function an aviation function was calculated using below given Eq. 5 taken from [89]:

$$O_j = X_k = \frac{1}{1 + e^{Net_j}}. \tag{5}$$

The Eq. 8 was used for receiving output and forwarding the input values through hidden layer in case of not using backpropagation method and not attaining output. For this, a sigmoid function is calculated by giving output back to the hidden layer and used feedforward concept for processing input. For predicting Zika virus, a Multilayer perceptron NN classifier was applied on zika dataset that gives the accuracy of 97%. Then, Lusk et al. have developed and trained ML algorithms namely SVM, LR, RF, KNN, GBT and DT and it was trained using data collected from January 2016-October 2017 [90]. The results show 90% of sensitivity for U.S. Zika Pregnancy and Infant Registry (USZPIR) dataset and 97% for Zika Active Pregnancy Surveillance System (ZAPSS).

Herry et al. have hypothesized that heart rate variability (HRV) in children suffering from ZIKV would yield a biomarker of fetal ZIKV exposure [91]. The main motive was to test the hypothesis in young children who are exposed to Zika virus during pregnancy. For evaluation they have used cubic SVM classifier that gives 92% of negative predictive value and 86% of positive predictive value. In the same year, Veiga et al. have also presented a ML algorithm that will be able to find ZIKV patients from structured and unstructured available data [92]. Their model was able to give the accuracy of 83% with textual information in medical records and 76% of accuracy was achieved on image reports and in classifying data without textual information. Dadheech et al. have found that RNA test was performed on the patients who were found to be infected with ZIKV [93]. The results obtained after applying multilayer perceptron and probabilistic hybrid optimization algorithm on training the specified characteristics show greater accuracy rate that is 91.25%.

5.7 Prediction of Sepsis

The sepsis outcomes get improved by rapid antibiotic administration but early diagnosis is still not an easy task because of complex presentation. So, a model using rapidly electronic health record (EHR) data was developed by Masino et al. that was capable of recognising infant sepsis at least 4 h before the its recognition clinically [32]. For their study they have considered infants retrospective control case who is hospitalized for more than 48 h in the Neonatal Intensive Care Unit (NICU) at the Children's Hospital of Philadelphia during the period of September 2014 to November 2017. The results give around 85–87% of accuracy with no significance differences. In this term, Song et al. examined the feasibility of prediction model using ML technology along with non-invasive vital sign data and tested it on electronic medical record data [33]. After examination, they have found for late onset neonatal sepsis algorithm the AUC of 48 h prediction model is 86.1% and it was 86.8% for onset detection model. One year later Late onset sepsis in preterm infant was predicted by Cabrera-Quiros et al. 2021 using multiple patient monitoring signals of 24 h prior to onset [34]. They have used number of features containing respiration, heart rate variability, body motion, difference of late onset sepsis and C of upto 5 h preceding cultures, antibiotics and resuscitation that gives accuracy of 79% and mean precision was 82%. Further, Kopanitsa et al. have considered ML prospective for prediction and risk management of sepsis using rela world evidence data for identification of sepsis risk in early stages of pregnancy [35]. They have tested their model on electronic health records of patients taken from perinatal Center of the Almazov perinatal medical center that gives AUC of 95%. In the same year, Hsu et al. have also used ML approaches for prediction of in-hospital mortality neonatal through data-driven learning [36]. The data considered by them was consists of total 1095 neonates suffered from sepsis in a tertiary-level NICU in Taiwan enrolled under the period of August 2017 to July 2020. The ML methods used for prediction work were DNN, KNN, SVM, RF and EGB. Out of all models, the best outcome was achieved using DNN that gives 92.3% of AUC, 95.64% of accuracy with 0.74, 95% CI 0.79–0.69 of cohen's kappa coeffieicnt value and 0.75, 95% CI 0.80–0.70 Mathews correlation coefficient value.

5.8 Prediction of Preeclampsia (PE) Hypertensive

Preeclampsia (PE) is a hypertensive complication affecting pregnancy and it can't be cured. However existing study shows that PE can be prevented by its early prediction. So, ANN model was developed by Tejera et al. for classification of women who have preclamptic and hypertensive pregnancy in different gestation ages using maternal heart rate variability indexes [94]. The results obtained using proposed model gives the sensitivity of 80% and specificity of 85–90% for PE that increases normal and hypertensive groups of pregnancy. Further neuro fuzzy ML approach was proposed by Moreira et al. for prediction of complex hypertensive disorder in pregnancy called HELLP syndrome [95]. Their proposed classifier serves as suggestion mechanism for cloud based applications that give 68.5% of precision, 70.5% of F-measure and 75.6% of recall value.

Later, to improve the prediction of risk associated with PE level in pregnant women Tahir et al. have used NN and DL algorithms [96]. For analysing the proposed algorithm, they have used 17 parameters of 1077 patient data of two Makassar hospitals and Haji General Surabaya Hospital from duration of December 12th 2017 until February 12th 2018. Then required features were selected using PSO and final outcome obtained from it gives 95.12% of accuracy. A year later, Bennett et al. have done prediction of PE using cost-sensitive DNN and considered the severe imbalance data with racial disparities [97]. For validation of proposed model, Rachel has used data sources representing US diverse minority populations including different databases. From results it has been depicted that the proposed model was able to achieve the accuracy of 72.4% in prediction of PE among minority populations.

6 Comparative Analysis

This section contains the comparison table of reviews done on work performed by various researchers in the field of prediction of congenital diseases [98–102]. We have covered maximum papers on Fetal health, Congenital heart disease, CMV, IUGR, Sepsis, Hypertensive and chromosomal

Table 3 Analysis of congenital diseases

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Haghpanahi et al. (2014)	IOP Science [52]	Fetal Health	4 set electrodes placed on ECG, mother abdomen datasets provided by Physio Net/CinC challenge 2013	Kalman filtering, Gaussian model selection	To validate their approach, there is a need to use a large dataset	SE = 98.2, PPV = 95.75
Abbasi et al. (2017)	IJNS [53]	Fetal Health	Preterm fetal sheep model cohort EEG recordings	Type-2 fuzzy logic system classifier	Need to validate model on other datasets	Average overall performance rate = 94
Miao et al. (2018)	IJACSA [54]	Fetal Health	2126 clinical instances CTG dataset, UCI ML repository dataset	DNN, Ensemble learning	The training period is large compared to standard DNN architecture making updated parameters in the network noisy	Accuracy = 88.02, Recall = 94.30, Precision = 85.01, F-Score = 85.08
Akbulut et al. (2018)	Elsevier [55]	Fetal Health	3 clinicians detailed evaluations and maternal questionnaire obtained dataset	Averaged perceptron, DT, DF, Decision jungle, SVM, Neural network, Locally deep SVM	There is a need to work on making the system more reliable and encrypting the communication between systems. Also, increase the number of elements in the dataset	Accuracy = 87.5 for a test performed on 16 users
Pisapia et al. (2018)	Jama Pdeiatrics [57]	Fetal Health	Collected from 2008 to 2014 fetal ventriculomegaly 253 patients	Image-based ML	There is need to incorporate the fetal MRI-based predictive model into user-friendly software	Discovery cohort: Accuracy = 82, Sensitivity = 80, Specificity = 84 Replication on cohort: Accuracy = 91, Sensitivity = 75, Specificity = 95
Haghpanahi et al. (2014)	IEEE [52]	Fetal Health	Simulated data	CNN	The quantization of permutation will make 40 k to 50 k classes and 28 fetal examples in each class. This will cause difficulties in training and also reduces the resolution of prediction	Average spatial prediction error: 7 mm
Feng et al. (2017)	IEEE [58]	Chromosome Anomalies	Illumina exome genotyping array dataset	convolutional neural network (CNN)	There is need to facilitate DS gene therapy studies and developments of genetic medicine	Average accuracy = 99.3

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Neocleous et al. (2017)	IEEE [61]	Chromosome Anomalies	Considered total 122,362 cases of euploid and 967 aneuploid	ANN	Need to develop a two-stage procedure for the short and long markers groups for screening first and second aneuploidy	For trisomy 21: detection rate = 100, For Trisomies 13 and 18, Turner and triploid syndromes: detection rate = 80
Somasundaram (2018)	Wiley [60]	Chromosome Anomalies	Number of chromosome datasets resampled patches of images	Convolutional neural network	The pre-processing approaches need to be improve	Accuracy = 98.7, AUC = 97.0
Neocleous et al. (2018)	Wiley [59]	Chromosome Anomalies	Training dataset containing 72,054 euploid pregnancies, 295 and 305 cases of T21 and Other chromosomal abnormalities respectively	Feed-forward ANN, cfDNA	Their method is not able to correctly classify both stages used in the proposed work	For T21: Sensitivity = 97.1, Specificity = 99.5
Sharma et al. (2018)	IEEE [106]	Chromosome Anomalies	Bioimage chromosome classification dataset	Deep CNN	N/A	Accuracy = 90.42
Qin et al. (2019)	IEEE [62]	Chromosome Anomalies	Constructed their dataset of 1909 cases of karyotyping	Deep CNN	Not able to recognize the gently bent chromosomes	Accuracy = 99.2
Al-Kharraz et al. (2020)	IEEE [64]	Chromosome Anomalies	Biomedical imaginary Laboratory dataset	YOLOv2 Convolutional neural network (CNN)	Need to improve accuracy by increasing the count of metaphase images in the dataset. Also, segment out the touched and overlapped chromosomes and detect structural abnormalities	Accuracy = 94.11
Jaganathan et al. (2019)	IJAS [63]		MSS dataset taken from Cneter of prenatal diagnosis	L1-norm based Support Vector Regression (L1-SVR)	Need to perform classification other using ML approaches and construct an effectual benchmark dataset for improving the accuracy	Accuracy = 96.78%, Sensitivity = 94.81%, Specificity = 98.50%, Precision = 96.43%, F1-Score = 96.42%
Nimitha et al. (2022)	AIP [66]	Chromosome anomalies	Biomedical Imaging Laboratory dataset	VGG16 based neural network for transfer learning	N/A	Accuracy = 98.6 Precision = 95.5
Li et al. (2017)	Medicine [37]	Congenital Heart Disease	Based on comprehensive epidemiological data, they have collected their won data	BPNN	Need to work on large sample size	Accuracy = 91, Specificity = 78, Sensitivity = 68

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Luo et al. (2018)	PLoS [68]	Congenital Heart Disease	33,831 live births having 78 cases of CHD	Weight SVM, weighted RF, LR	They have tested the proposed model on survey data instead of diagnostic data	TPR and TNR of the three classifiers: above 65 and 93
Du et al. (2020)	IEEE [39]	Congenital Heart Disease	68,969 child patients ECG available at GZFEZX	Deep neural network	Need to further improve their model to work on unbalanced datasets classification	Accuracy = 92.3, Sensitivity = 74.73, Specificity = 94.07
Yoon et al. (2020)	Springer [38]	Congenital Heart Disease	child vitals database	KNN, SVM, and Logistic Regression	Need to collect more data that can be used for training ML algorithms. Also, need to test more ML algorithms on a dataset for getting a more accurate prediction outcome	Prediction accuracy using SVM = 73
Arnaout et al. (2021)	Nature Medicine [41]	Congenital Heart Disease	Five datasets: FETAL 125, OB-125, OB-4000, BCH-400, TWINS-10	Ensemble neural network	They have to refine the used technique for larger populations so that the same approach can be applied to other medical imaging diagnostic challenges	AUC = 99, Sensitivity = 95, Specificity = 95
Nurmaini et al. (2021)	MDPI [42]	Congenital Heart Disease	Microsoft common objects in context (COCO)	Mask-Fully convolutional neural network	Need to refine their model in larger populations	Intra-patients variation: Precision = 98.3 Inter-patient variation = 82.42
Ammarah et al. (2021)	IEEE [43]	Congenital Heart Disease	Lahore Hospital Pakistan samples of 518 mothers	KNN, DT, RF, SVM, deep learning model multilayer perceptron (MLP)	N/A	MLP classifier: Accuracy = 92.9
Truong et al. (2021)	Hindawi [44]	Congenital Heart Disease	UCI Machine Learning Heart Disease	SVM, KNN, RF, Decision tree	The sample size of dataset is small	Using deep learning: Accuracy = 94.2
Qu et al. (2022)	Frontier Science [45]	Congenital Heart Disease (CHD)	Retrospective cross-sectional study from northwest China dataset	Explainable Boosting Machine (EBM)	They have considered limited count of extreme levels exposed cases	Accuracy = 76, Sensitivity = 74, Specificity = 65
Buscema, et al. (2015)	PLOS One [70]	Intrauterine growth restriction (IUGR)	Existing study S1 dataset	Artificial Neural Networks (ANNs), K-Mean clustering	Need to discover each dataset variable prototypical features	Perform well

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Wosiak et al. (2016)	ACSIS [71]	Intrauterine growth restriction (IUGR)		Bagging with NB, KNN, C4.5 and SMO	Not able to perform well in case of unlabelled data	Hybrid accuracy = 80.88, Sensitivity = 80.9, Precision = 81.2
Bahado-Singh et al. (2019)	PLoS One [72]	Intrauterine growth restriction (IUGR)	Metabolites	Learning vector quantization (LVQ), SVM classifier	They have used a small sample size and also not validated their findings using an independent cohort	Accuracy = 91, Sensitivity = 83, Specificity = 80
Sufriyana et al. (2020)	Elsevier [73]	Intrauterine growth restriction (IUGR)	Prospective cohort study public dataset	Matthew correlation coefficient (MCC)	Not able to give good outcome in second and third trimester	Specificity = 90, Sensitivity = 95,
Pini et al. (2021)	Frontier Science [74]	Intrauterine growth restriction (IUGR)	Recording of Antepartum FHR that was collected at Azienda Ospedaliera Universitaria—Federico II (Napoli, Italy)	Radial basis function (RBF)-SVM,	Need to introduce additional features like ones that are inspired to fractal and multiscale analysis that help in improving the accuracy of classification. Also, need to investigate other ML and DL approaches along with validation of existing approaches on external datasets	Accuracy = 93, Sensitivity = 93, Specificity = 84
Crockart et al. (2021)	Elsevier [75]	Intrauterine growth restriction (IUGR)	Week20_ONLY and T3 dataset	Stochastic gradient decent, KNN, RF, Logistic Regression	The data used in the proposed work may become a limiting factor for further analysis so there is a ned to further develop an existing model	Accuracy = 93, Recall = 91, F1-Score = 95
Teng et al. (2022)	Springer [76]	Intrauterine growth restriction (IUGR)	Doppler velocimetry and ultrasound biometry of second trimester	SVM	Need to consider other important parameters	Abdominal circumference = 11, Uterine RI = 15, Uterine PI = 8
Aslam et al. (2022)	MDPI [77]	Intrauterine growth restriction (IUGR)	Two datasets: first was introduced by [26] and the second one by [24]	RF, SVM, KNN, Gradient Boosting (GB)	Need to investigate their model on huge dataset	Using RF: Accuracy = 97, F1-Score = 98

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Arav-Boger et al. (2008)	Bio Information Biological Insight [81]	CMV	Two datasets: 54 women who contracted acute CMV infection during their pregnancy amniotic fluids obtained 54 sequences of CMV gene UL144 and the second one is 80 sequences of 4 genes US28, UL144, UL146 and UL147 taken from urine, saliva and blood of 20 infected infants	Artificial Neural networks	They do not have oseqquence imformation on gN and gB hypervariable CMV genes	Predicting outcome for US28, UL147 = 80, UL144, UL146 = 85
Rogers et al. (2020)	Springer [83]	Cytomegalovirus (CMV)	They have portrayed several CMV images along with different views taken from blood vessels, fluids, and saliva from the body fluids, blood vessels, saliva	Hierarchical ranking CNN (HR-CNN), K-means clustering, High SVM (HSVM)	Need to improve the feature selection approach to get a better outcome	Accuracy = 91.17 Error = 8.823, Sensitivity = 92.82, Specificity = 97.47, Precisiom = 86.66, F1-Score = 88.86, Kappa = 76.47, Geometric mean = 85.12, Mathews correlation Coefficient = 85.83
Deepa et al. (2020)	ICTACT [84]	Cytomegalovirus (CMV)	CyTOF	Deep CNN	To confirm further results, they need to establish disease-specific laboratory tests for the diagnosis of CMV infection	AUPRC = 91
Lee et al. (2021)	Scientific Reports [86]	Cytomegalovirus (CMV)	456 CMV esophagitis and 666 endoscopic images	Logistic regression, RF	They have not tested their system for other clinical diagnosis applications	Sensitivity = 100, Specificity = 100, Positive predictive value = 100,Negative predictive value = 100, Accuracy = 100, Area under the receiver = 1
Eisenberg et al. (2014)	IOP Science [87]	Cytomegalovirus (CMV)	Taken from 4 set electrodes placed on ECG, mother abdomen datasets. This is provided by PhysioNet/CinC challenge 2013	Kalman filtering, Gaussian model selection	To validate their approach, there is a need to use a large dataset	SE = 98.2, PPV = 95.75

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Jiang et al. (2018)	Elsevier [88]	Zika Virus	Multidisciplinary datasets	backward propagation neural network (BPNN), GBM, RF	There are mismatches in spatial and temporal scales between presence points and predictor variables that may affect the prediction of ZIKV transmission risk level	BPNN:AUC:96.6, GBM: AUC = 96.4, RF: 96.3
Mahalakshmi et al. (2019)	IJRTE [89]	Zika Virus	They have created a dataset based on the collected information	Multilayer Perceptron neural network (MLPNN) using cloud	Not able to give a final prediction as there is a need to go through RNA test for confirmation	Accuracy = 97.9, Sensitivity = 98, Specificity = 97
Lusk et al. (2020)	Wiley [90]	Zika virus	USZPIR, ZAPSS	VM, RF, logistic regression, and GBT	The proposed ML methods are not able to classify individual types of Zika-related birth defects. Also, the methods are designed specifically for USZPIR and ZAPSS surveillance systems	USZPIR: Sensitivity = 96, ZAPSS: Sensitivity = 97
Herry et al. (2021)	Physiology Measure [91]	Zika Virus	4–25 months 21 children HRV properties from Brazil	Cubic support vector classifier	N/A	Negative prediction value = 92, Positive predictive value-86
Veiga et al. (2021)	Scientific Reports Scientific Reports [92]	Zika Virus	They have built it through linkage of SINASC and RESP containing different information related to birth	KNN, RF, Gradient Boosting (GB)	They have not validated their approach on different open-access datasets	Textual information in image reports and medical records: Accuracy = 83, Without textual information = 76
Dadheech et al. (2022)	Hindawi [93]	Zika Virus	Operational data and production environment dataset	Multilayer perceptron with a probabilistic optimization strategy	Not able to get a real-time dataset	Accuracy = 91.25 with minimal timing of 0.15 s for cryptosystem processing
Masino et al. (2019)	PLOS One [32]	Sepsis	Automatically extracted from EHR and anonymized before transferring to study database	LR, NB, SVM with radial basis function kernel, KNN, RF, AdaBoost, and gradient boosting	Need to introduce more input features to improve the performance of the prediction	AUC = between 85–87 with no significant differences

Table 3 (continued)

Authors	Publication	Type of disease	Dataset	Technique	Limitations	Outcomes
Song et al. (2020)	Elsevier [33]	Sepsis	MIMIC-III database	Logistic regression (LR), Gaussian Naïve Bayes (GNV), Decision tree, Gradient boosting, Adaptive boosting, Bagging classifier, RF and Multilayer perceptron	Need to conduct external validation on various datasets along with doing the clinical verification of the proposed model	48-h prediction model: 86.1, Onset detection model: 86.8
Cabrera-Quiros et al. (2021)	Elsevier [34]	Sepsis	Almazov perinatal medical center electronic health records located in Saint-Petersburg, Russia	GBR, RF, LR, and Voting regression	There is need to facilitate early identification of risk factors	Accuracy = 95
Kopanitsa et al. (2021)	MDPI [35]	Sepsis	1095 Neonates experienced sepsis in teritiary level in Taiwan from period of 2017 to 2020	DNN, KNN, SVM, RF, EGB	They have not investigated the real time risk mortality adjustment for further optimizing the treatment along with improving the NICU outcomes	Best accuracy using DNN = 95.64
Tejera et al. (2011)	MDPI [94]	Hypertensive complication	568 ECG records	ANN	NA	Sensitivity = 80%, Specificity = vary from 85–90%
Moreira et al. (2018)	Wiley [95]	Hypertensive complication	205 parturient women diagnosed with hypertensive disorder	ANN, ANFIS, Fuzzy logic	Need to add other approaches for classification and clustering data to get a better outcome	AUC = 82.9%
Tahir et al. (2018)	Emitter [96]	Hypertensive complication	Original data	Particle swarm optimization, Deep learning, NN	N/A	Accuracy = 95.12%
Bennett et al. (2022)	PLoS [97] ONE	Hypertensive complication	Exas Public Use Data Files (PUDF), Magee Obstetric Medical and Infant (MOMI), and Oklahoma PUDF	Cost-sensitive DNN (CSDNN), Bayesian optimization	Further, there is a need to work on both early and late PE. Also, need to explore data for minority groups and improve the outcome	Using CSDNN: AUC = 62.3%, Using DNN for Oklahoma PUDF database: AUC = 58% MOMI database: AUC = 72.4%

anomalies given in Table 3. The motive behind it is to cover the various ML and DL techniques used for the prediction of congenital diseases along with limitations of individual work and dataset used for it.

From the above Table 3 comparison, we have found that several congenital diseases are affecting the fetus and infants. In existing work, researchers have used different ML or DL classification algorithms and applied them to various datasets to show the performance of their models in terms of different parameters [103, 104]. They can achieve good outcomes using different classifiers but still, every model has its limitations which are given in the last column of the table. We have also specified different metrics used for evaluation purpose that help further researchers in comparing and evaluating their work in the same field [105].

7 Discussion

The primary goal of this research is to give a review on current research on prediction of congenital diseases. Furthermore, we addressed various papers related to different congenital diseases published in different years [107, 108]. Along with this, we tried to cover ML and DL approaches used to predict such diseases, dataset used by individual, outcomes and metrics used for testing the models. As a result, this article has provided a compilation of valuable results with sufficient precision, accuracy, efficiency and F-score.

7.1 Investigation 1: Thorough Investigated Year by Year Evaluation of Congenital Disease Prediction Studies?

There are number of congenital diseases that exist in the universe. We performed an efficient search for selection of research papers on Google scholar, IEEE Xplore, Research gate and science direct. Some of the papers were available openly and few need to purchase from sites to get full access of the paper [109, 110]. The Fig. 7 show the distribution of the chosen research and review articles by year between 2008 and 2022 taken from different databases, Journals and Conferences.

The highest number of research articles is 21 that have been reported in the year 2018, while in the year 0f 2008 and 2010 the lowest count of technical papers was found [111]. Figure 7 depicts the distribution of various diseases according to the count of papers consider for individual disease by focusing on the prediction of congenital disease using ML and DL methods. The most thoroughly studied congenital diseases papers as well as year by year overview of such papers is given in below Fig. 8.

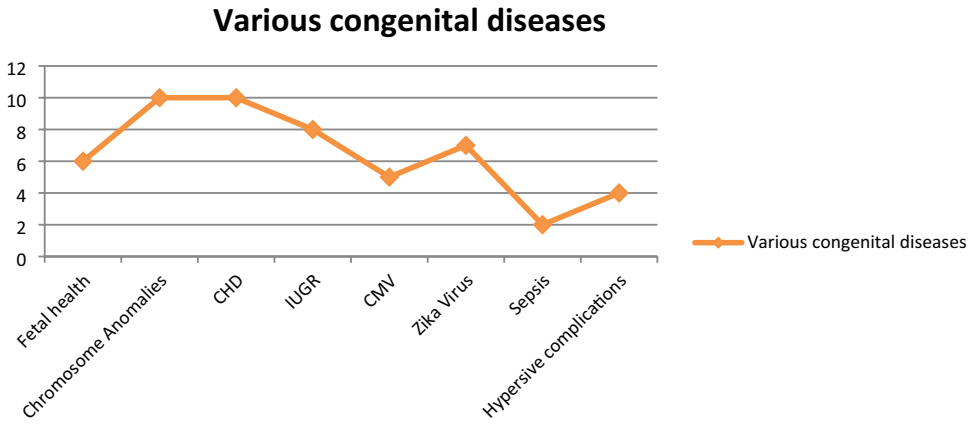

Fig. 7 Various congenital diseases papers contributed in the paper

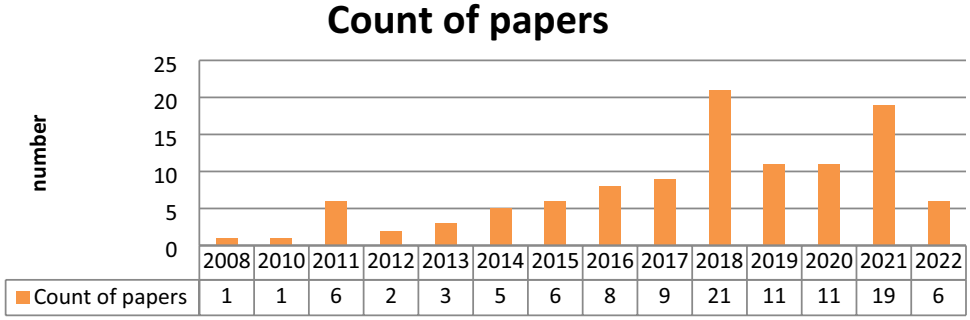

Fig. 8 Year wise distribution of papers

7.2 Investigation 2: Which ML and DL Techniques Have Been Used for Prediction of Congenital Diseases?

There are number of ML and DL algorithms that have made contributions in the body of knowledge and in the field of diseases research [112–114]. The research work mentioned in the literature has focused on use of such techniques in the prediction of congenital diseases shown in Fig. 9. The papers used Type 2 fuzzy logic system classifier (36,75, 88), DNN (37, 44, 45, 51, 53, 67, 72, 76, 85, 89, 90), Ensemble learning (37, 55, 59, 67, 76, 80, 83, 84, 85), Decision tree (17, 57, 58), SVM (17, 46, 50, 52, 57, 58, 61, 64, 66, 67, 71, 79, 85), Neural network (17), CNN (39, 43, 44, 45, 47, 40, 56, 71, 72), ANN(42, 41, 49, 60, 67, 77, 87, 88), Random forest (28, 50, 57, 58, 65, 67, 74, 76, 78, 80, 82, 83, 84, 85), Logistic regression (28, 46, 50, 5, 65, 74, 78, 82, 83), Radial basic function (37, 64), KNN(52, 57, 58, 61, 65, 67, 80, 82, 85), K-means clustering (42, 60, 71), Naïve Bayes (61, 82, 83), Linear regression(84) have been the most often employed learning models for prediction of congenital diseases.

7.3 Investigation 3: What are the Limitations of Existing Congenital Disease Prediction Models Using ML and DL Techniques?

ML is a generic approach and to construct a situation to accurately get the real outcome there is a need for mathematical equations. Such equations represent the truth that simplifies the problem [115, 116]. So, there is a need to develop AI model that can predict the real outcome of congenital in a better way. As there is no such generic solution so every problem needs to be treated accordingly [117–119]. In relation to it, there is a need to add some mathematical formulas for the detection of congenital health defects to improve the prediction rate. In existing AI models, there is a need to improve the accuracy by increasing the count of metaphase images in the dataset but some problems took place. Overfitting and underfitting are two main errors that restrict AI learning-based models to get an improved outcome. Overfitting takes place when a noise comes in the training dataset that reduces the performance of the models of a new dataset. On other hand, a new dataset can't be trained or generalized in underfitting which affects the performance of the systems [120]. So, AI models need to be trained in a way that they will be able to give an accurate performance in case of both balanced and unbalanced datasets.

Regarding features, most of the AI models were not developed to classify individual types of congenital-associated birth defects. Also, there is a need to consider additional features other than those used in existing work that helps in improving the accuracy of classification [121, 122]. As the selected features play an important role so more work is required on improving feature selection approaches to get a better outcome of classification. The AI-based models are easy to implement but at the same time, it is very critical to get an accurate outcome from them.

7.4 Investigation 4: What Metrics are Used for Evaluating ML and DL Models?

Prediction models evaluation is vital in finding their metrics and performance that assist in prediction of accuracy and comparing various models [123]. The summary of evaluation metrics is given in the Table 4 that was used by different researchers to evaluate their prediction models in predicting congenital diseases. TP stands for true positive that denotes the cases recognized correctly, FP stands for false positive that denotes cases identified negatively and incorrectly by different techniques, TN denotes the true negative which tells the outcomes that are identified correctly as negative by the technique, FN is the false negative that denotes

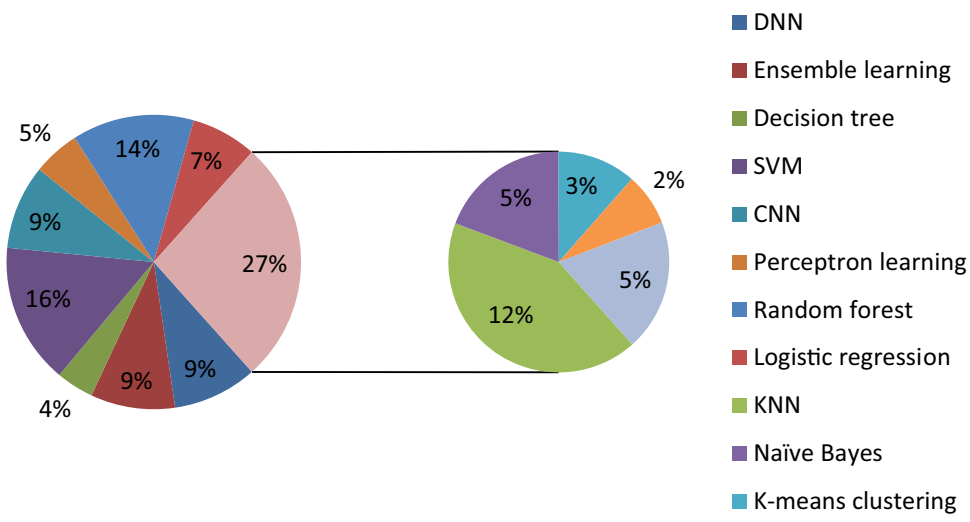

Fig. 9 Congenital disease prediction methods

Table 4 Evaluation parameters

Metrics	Description	Formula
Acc	It evaluate the overall benchmark performance	$= \frac{\sum_{i=1}^{l} TP_{i+} \sum_{i=1}^{l} TN_i}{\sum_{i=1}^{l} TP_{i+} \sum_{i=1}^{l} TN_{i+} \sum_{i=1}^{l} FP_{i+} \sum_{i=1}^{l} FN_i}$ (6)
Sensitivity/recall (Re)	It tells the number of identified real instances among all correct cases	$= \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} TP_i + \sum_{i=1}^{l} FN_i}$ (7)
Specificity	It quantifies actual negative observations proportion compared to total predicted negative observations by an algorithm	$= \frac{\sum_{i=1}^{l} TN_i}{\sum_{i=1}^{l} TN_{i+} \sum_{i=1}^{l} FP_i}$ (8)
Precision (Pr)	It is used to identify real instances from all predicted cases	$= \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} TP_i + \sum_{i=1}^{l} FP_i}$ (9)
F1-score	It is calculated using measurements of weighted harmonic mean of the accuracy and recall	$= 2 \times ((Pr \times Re)/(Pr + Re))$ (10)
Area under curve (AUC)	It is used in case of binary images	–

negative cases predicted incorrectly by the model. In the reported work, performance of classification was identified using different parameters namely accuracy (Acc), Precision (Pr), Area under the curve (AUC), Sensitivity/Recall (Re), F1-score (F) and Specificity (Sp) [124].

8 Conclusion

The review done in this study aim to summarize different research directions for ML or DL based prediction of congenital diseases. In the field of healthcare significance has been marked by AI. This paper presents critical and analytical examination of the ML and DL used in early prediction of congenital diseases namely CHD, CMV, Chromosome anomaly, Sepsis, Hypertensive complications, Zika virus and IUGR using medical imaging or numerical data. For study, we have shortlisted one hundred fifteen papers from year between 2008 and 2022 of different online digital libraries such as IEEE Xplore, Science direct, Google scholar and Elsevier. In the first section of the paper, introduction and organization of the paper is covered followed by Research methodology used for conducted the review on the topic that consists of quality assessment and investigations considered while conducting the survey. We have also mentioned some inclusion and exclusion standards used for selecting and rejecting papers. Later, framework for predicting congenital disease is covered that give general overview of different steps that need to be performed while predicting different congenital diseases using ML or DL models. Then, number of congenital disease or disorder is covered in tabular form that covers illness caused by such diseases along with causes of the same. The reported work on number of congenital diseases along with its comparative analysis using different techniques, dataset and parameters such as accuracy, specificity, sensitivity, F-score and AUC has been portrayed thoroughly. From this review, we have concluded that most of the existing literature works employed ML and DL techniques especially SVM, RF and CNN. To recapitulate, the use of AI for predicting congenital diseases is better understand by medical experts that lead to more better proposal for future work. Even there is huge advancements has been seen over the past several years, the area of accurate congenital diseases diagnosis faces number of limitations that need to be resolved and accordingly improving the constantly for treating emerging diseases effectively.

References

1. Waldorf KMA, McAdams RM (2013) Influence of infection during pregnancy on fetal development. Reproduction (Cambridge, England). https://doi.org/10.1530/REP-13-0232
2. McAuley JB (2011) Congenital toxoplasmosis. J Pediat Infect Dis Soc. https://doi.org/10.1093/jpids/piu077
3. Spitzer MH, Nolan GP (2016) Mass cytometry: single cells, many features. Cell 165:780–791. https://doi.org/10.1016/j.cell.2016.04.019
4. https://www.atsu.edu/faculty/chamberlain/Website/lectures/lecture/congen.htm
5. Tejera E, Areias MJ, Rodrigues A, Ramõa A, Nieto-Villar JM, Rebelo I (2011) Artificial neural network for normal, hypertensive, and preeclamptic pregnancy classification using maternal heart rate variability indexes. J Matern Fetal Neonatal Med. https://doi.org/10.3109/14767058.2010.545916
6. WHO fact sheet on congenital anomalies, updated September 2016, Available at www.who.int/mediacentre/factsheets/fs370/en/, Accessed 5 Oct 2018
7. Shahid N, Rappon T, Berta W (2019) Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS ONE. https://doi.org/10.1371/journal.pone.0212356
8. Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. https://doi.org/10.4236/jilsa.2017.91001
9. Workowski KA, Bachmann LH, Chan PA, Johnston CM, et al. (2021) Sexually Transmitted Infections Treatment Guidelines. MMWR Recommendation Report, 181–187
10. Workowski KA, Berman S (2010) Sexually transmitted diseases treatment guidelines. Morbidity and Mortality Weekly Report, 107–113

11. Liu W, Yu Z, Raj B, Yi L, Zou X, Li M (2015) Efficient autism spectrum disorder prediction with eye movement: a machine learning framework. Int Conf Affect Comput Intell Interact (ACII). https://doi.org/10.1109/ACII.2015.7344638
12. Marty FM, Ljungman P, Chemaly PF, Maertens J, Dadwal SS, Duarte RF, Haider S et al (2017) Letermovir prophylaxis for cytomegalovirus in hematopoietic-cell transplantation. N Engl J Med. https://doi.org/10.1056/NEJMoa1706640
13. Santis MD, Luca CD, Mappa I, Spagnuolo T et al (2012) Syphilis infection during pregnancy: fetal risks and clinical management. Infect Dis Obstert Gynecol. https://doi.org/10.1155/2012/430585
14. Wu X, Long E, Lin H, Liu Y (2016) Prevalence and epidemiological characteristics of con-genital cataract: a systematic review and meta-analysis. Sci Rep. https://doi.org/10.1038/srep28564
15. Veiga RV, Schuler-Faccini L, França GVA, Andrade RFS et al (2021) Classification algorithm for congenital Zika syndrome: characterizations, diagnosis and validation. Sci Rep. https://doi.org/10.1038/s41598-021-86361-5
16. Stevens GA, White RA, Flaxman SR, Price H, Jonas JB, Keeffe J, Leasher J et al (2013) Global prevalence of vision impairment and blindness: magnitude and temporal trends. Ophthalmology. https://doi.org/10.1016/j.ophtha.2013.05.025
17. Taruscio D, Mantovani A, Carbone P, Barisic I, Bianchi F et al (2015) Primary prevention of congenital anomalies: recommendable, feasible and achievable. Public Health Genom. https://doi.org/10.1159/000379739
18. Tejera E, Areias MJ, Rodrigues A, Ramoa A, Nieto-Villar JM, Rebelo I (2011) Artificial neural network for normal, hypertensive, and preeclamptic pregnancy classification using maternal heart rate variability indexes. J Matern Fetal Neonatal Med. https://doi.org/10.3109/14767058.2010.545916
19. Jung KH, Choi J, Gong EJ, Lee JH, Choi KD et al (2019) Can endoscopists diferentiate cytomegalovirus esophagitis from herpes simplex virus esophagitis based on gross endoscopic fndings? Medicine. https://doi.org/10.1097/MD.0000000000015845
20. Uchechi HO, Ifeyinwa OM, Asemota E, Okpokam D (2018) Seroprevalence of transfusion-transmissible infections (hbv, hcv, syphilis and hiv) among prospective blood donors in a tertiary health care facility in Calabar Nigeria; an eleven years evaluation. BMC Public Health. https://doi.org/10.1186/s12889-018-5555-x
21. Lawton B, MacDonald EJ, Brown SA, Wilson L, Stanley J, Tait JD (2014) Preventability of severe acute maternal morbidity. Am J Obstet Gynecol. https://doi.org/10.1016/j.ajog.2013.12.032
22. Huanga Y, Wanga H, Li T, Li C, Tanga J et al (2021) Comparison of detection strategies for screening and confirming congenital cytomegalovirus infection in newborns in a highly seroprevalent population: a mother-child cohort study. Lancet Region Health West Pacific. https://doi.org/10.1016/j.lanwpc.2021.100182
23. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR (2016) Zika virus and birth defects-reviewing the evidence for causality. N Engl J Med. https://doi.org/10.1056/NEJMsr1604338
24. Adachi KN, Saines KN, Klausner JD (2021) *Chlamydia trachomatis* screening and treatment in pregnancy to reduce adverse pregnancy and neonatal outcomes: a review. Front Public Health. https://doi.org/10.3389/fpubh.2021.531073
25. Wallace J, Pitts M, Liu C, Lin V, Hajarizadeh B, Richmond J, Locarnini S (2017) More than a virus: a qualitative study of the social implications of hepatitis b infection in china. Int J Equity Health. https://doi.org/10.1186/s12939-017-0637-4
26. Oliveira D, W. K. et al. (2016) Increase in reported prevalence of microcephaly in infants born to women living in areas with confrmed Zika virus transmission during the frst trimester of pregnancy-Brazil. Morbidity and Mortality Weekly Report. https://doi.org/10.15585/mmwr.mm6509e2
27. Franca GVA, Faccini LS, Oliveira WK, Henriques CMP et al (2016) Congenital Zika virus syndrome in Brazil: a case series of the frst 1501 livebirths with complete investigation. Lancet. https://doi.org/10.1016/S0140-6736(16)30902-3
28. Aragao MDFV, Linden VVD, Brainer-Lima AM et al (2016) Clinical features and neuroimaging (CT and MRI) fndings in presumed Zika virus related congenital infection and microcephaly: retrospective case series study. BMJ Clin Res. https://doi.org/10.1136/bmj.i1901
29. Ciobanu A, Khan N, Syngelaki A, Akolekar R, Nicolaides KH (2019) Routine ultrasound at 32 vs 36 weeks' gestation: prediction of small-for-gestational-age neonates. Ultrasound Obstet Gynecol. https://doi.org/10.1002/uog.20258
30. Gaccioli F, Aye ILMH, Sovio U, Charnock-Jones DS, Smith GCS (2018) Screening for fetal growth restriction using fetal biometry combined with maternal biomarkers. Am J Obstet Gynecol. https://doi.org/10.1016/j.ajog.2017.12.002
31. Bahado-Singh RO, Yilmaz A, Bisgin H, Turkoglu O, Kumar P et al (2019) Artificial intelligence and the analysis of multiplatform metabolomics data for the detection of intrauterine growth restriction. PLoS One. https://doi.org/10.1371/journal.pone.0214121
32. Masino AJ, Harris MC, Forsyth D, Ostapenko S, Srinivasan L et al (2019) "Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PLoS One. https://doi.org/10.1371/journal.pone.0212665
33. Song W, Jung S, Baek H, Choi C, Jung Y, Yoo S (2020) Development of prediction model for the early detection of late-onset neonatal sepsis using machine learning. JMIR Med Inform. https://doi.org/10.2196/15965
34. Quiros LC, Kommers D, Wolvers MK, Oosterwijk L, Arents N et al (2021) Prediction of late-onset sepsis in preterm infants using monitoring signals and machine learning. Crit Care Explor. https://doi.org/10.1097/CCE.0000000000000302
35. Kopanitsa G, Metsker O, Paskoshev D, Greschischeva S (2021) Identification of risk factors and prediction of sepsis in pregnancy using machine learning methods. Procedia Comput Sci. https://doi.org/10.1016/j.procs.2021.10.040
36. Hsu JF, Chang YF, Cheng HJ, Yang C, Lin CY, Chu SM et al (2021) Machine learning approaches to predict in-hospital mortality among neonates with clinically suspected sepsis in the neonatal intensive care unit. J Personal Medicien, MDPI. https://doi.org/10.3390/jpm11080695
37. Li H, Luo M, Zheng J, Luo J, Zeng R, Feng N, Du Q, Fang J (2017) An artificial neural network prediction model of congenital heart disease based on risk factors. Medicine. https://doi.org/10.1097/MD.0000000000006090
38. Yoon SA, Hong WH, Cho H (2020) Congenital heart disease diagnosed with echocardiogram in newborns with asymptomatic cardiac murmurs: a systematic review. BMC Pediator. https://doi.org/10.1186/s12887-020-02212-8
39. Du Y, Huang S, Huang C, Maalla A, Lia H (2020) Recognition of child congenital heart disease using electrocardiogram based on residual of residual network. IEEE Int Conf Progress Inform Comput (PIC). https://doi.org/10.1109/pic50277.2020.9350802
40. Meda JT, Mushiri T (2020) Predicting Congenital Heart Diseases Using Machine Learning. Proceedings of the 2nd African International Conference on Industrial Engineering and Operations Management, pp. 1716–1725.
41. Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ (2021) An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nat Med. https://doi.org/10.1038/s41591-021-01342-5
42. Nurmaini S, Rachmatullah MN, Sapitri AI, Darmawahyuni A et al (2021) Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for

congenital heart defects detection. Artif Intell-Based Appl Med Imag. https://doi.org/10.3390/s21238007
43. Ammarah UE, Bukhari F, Idrees M, Iqbal W (2021) Predictive analysis of congenital heart defects prior to birth. Int Conf Robot Automat Ind (ICRAI). https://doi.org/10.1109/ICRAI54018.2021.9651436
44. Truong VT, Nguyen BP, Nguyen-Vo TH, Mazur W, Chung ES et al (2022) Application of machine learning in screening for congenital heart diseases using fetal echocardiography. Int J Cardiovasc Imaging 38:1007–1015
45. Qu Y, Deng X, Lin S, Han F, Chang HH, Ou Y, Nie Z et al (2022) Using innovative machine learning methods to screen and identify predictors of congenital heart diseases. Front Cardiovasc Med. https://doi.org/10.3389/fcvm.2021.797002
46. Say L, Chou D, Gemmill A, Tunçalp O, Moller AB, Daniels J et al (2014) Global causes of maternal death: a WHO systematic analysis. Lancet Glob Health. https://doi.org/10.1016/S2214-109X(14)70227-X
47. Abu-Raya B, Maertens K, Edwards KM, Omer SB, Englund JA, Flanagan KL (2020) Global perspectives on immunization during pregnancy and priorities for future research and development: an international consensus statement. Front Immunol. https://doi.org/10.3389/fimmu.2020.01282
48. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP (2018) Machine learning in cardiovascular machine: are we there yet? Heart (Bristish Cardiac Society). https://doi.org/10.1136/heartjnl-2017-311198
49. Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A, Yanac TG (2021) Multiple disease prediction using machine learning algorithms. Materials Today Proc. https://doi.org/10.1016/j.matpr.2021.07.361
50. Lin KH, Hu YJ (2018) Application of machine learning to immune disease prediction. Int J Eng Innov Technol. https://doi.org/10.1038/s41746-020-0229-3
51 Li H, Luo M, Zheng J, Luo J, Zeng R, Feng N, Du Q, Fang J (2017) An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. Medicine (Baltimore). https://doi.org/10.1097/MD.0000000000006090
52. Haghpanahi M, Borkholder DA (2014) Fetal QRS extraction from abdominal recordings via model-based signal processing and intelligent signal merging. Physiol Meas. https://doi.org/10.1088/0967-3334/35/8/1591
53. Abbasi H, Bennet L, Gunn AJ, Unsworth CP (2017) Robust wavelet stabilized "footprints of uncertainty" for fuzzy system classifiers to automatically detect sharp waves in the EEG after hypoxia ischemia. Int J Neural Syst. https://doi.org/10.1142/S0129065716500519
54. Miao JH, Miao KH (2018) Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification. International Journal of Advanced Computer Science and Application. https://doi.org/10.14569/IJACSA.2018.090501
55. Akbulut A, Ertugrul E, Topcu V (2018) Fetal health status prediction based on maternal clinical history using machine learning techniques. Computer Methods Progr Biomed. https://doi.org/10.1016/j.cmpb.2018.06.010
56. Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B, Sun J (2014) PARAMO: a PARAllel predictive modeling platform for healthcare analytic research using electronic health records. J Biomed Inform. https://doi.org/10.1016/j.jbi.2013.12.012
57. Pisapia JM, Akbari H, Rozycki M, Goldstein H, Bakas S, Rathore S et al (2018) Use of fetal magnetic resonance image analysis and machine learning to predict the need for postnatal cerebrospinal fluid diversion in fetal ventriculomegaly. JAMA Pediatr. https://doi.org/10.1001/jamapediatrics.2017.3993
58. Feng B, Samuel DC, Hoskins W, Guo Y, Zhang Y, Tang J, Meng Z (2017) Down syndrome prediction/screening model based on deep learning and illumina genotyping array. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 347–352. https://doi.org/10.1109/BIBM.2017.8217674
59. Neocleous AC, Nicolaides KH, Schizas CN (2018) Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems. Ultrasound Obstet Gynecol. https://doi.org/10.1002/uog.17558
60. Somasundaram D (2018) Machine learning approach for homolog chromosome Classification. Int J Imaging Syst Technol. https://doi.org/10.1002/ima.22287
61. Neocleous AC, Nicolaides KH, Schizas CN (2017) Intelligent noninvasive diagnosis of aneuploidy: raw values and highly imbalanced dataset. IEEE J Biomed Health Informatics. https://doi.org/10.1109/JBHI.2016.2608859
62. Qin Y, Wen J, Zheng H, Huang X, Yang J, Song N, Zhu YM, et al. (2019) Varifocal-net: a chromosome classification approach using deep convolutional networks. Computer vision and pattern recognition IEEE TMI for future publication. https://doi.org/10.1109/TMI.2019.2905841
63. Jaganathan M, Gopal R, Kiruthika VR (2019) Modelling an effectual feature selection approach for predicting down syndrome using machine learning approaches. International Journal of Aquatic Science, pp. 1238–1249
64. Al-Kharraz MS, Elrefaei LA, Fadel MA (2020) Automated system for chromosome karyotyping to recognize the most common numerical abnormalities using deep learning. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3019937
65. Xu X, Wang L, Cheng X, Ke W, Jie S, Lin S, Lai M, Zhang L, Li Z (2022) Machine learning-based evaluation of application value of the USM combined with NIPT in the diagnosis of fetal chromosomal abnormalities. Math Biosci Eng. https://doi.org/10.3934/mbe.2022197
66. Nimitha N, Abbiraamavallee S, Elakiya E, Harini J, Kotishree V (2022) Supervised chromosomal anomaly detection using VGG-16 CNN model. AIP Conf Proc. https://doi.org/10.1063/5007249 1
67. Bhardwaj P, Bhandari G, Kumar Y, Gupta S (2022) An investigational approach for the prediction of gastric cancer using artificial intelligence techniques: a systematic review. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-022-09737-4
68. Luo Y, Li Z, Guo H, Cao H, Song C, Guo X, Zhang Y (2017) Predicting congenital heart defects: a comparison of three data mining methods. PLoS ONE. https://doi.org/10.1371/journal.pone.0177811
69. Gupta A, Koul A, Kumar Y (2022) Pancreatic Cancer Detection using Machine and Deep Learning Techniques. 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2, 151–155. https://doi.org/10.1109/ICIPTM54933.2022.9754010
70. Buscema M, Grossi E, Montanini L, Street ME (2015) Data mining of determinants of intrauterine growth retardation revisited using novel algorithms generating semantic maps and prototypical discriminating variable profiles. PLoS ONE. https://doi.org/10.1371/journal.pone.0126020
71. Wosiak A, Zamecznik A, Jarosik KN (2016) Supervised and unsupervised machine learning for improved identification of intrauterine growth restriction types. Proceedings of the federated conference on computer science and information systems, https://doi.org/10.15439/2016F515
72. Bahado-Singh RO, Yilmaz A, Bisgin H, Turkoglu O, Kumar P et al (2019) Artificial intelligence and the analysis of multiplatform metabolomics data for the detection of intrauterine growth restriction. Metabol Anal Intrauter Growth Restrict. https://doi.org/10.1371/journal.pone.0214121

73. Sufriyana H, Wu YW, Su YCU (2020) Prediction of preeclampsia and intrauterine growth restriction: development of machine learning models on a prospective cohort. JMIR Med Inform. https://doi.org/10.2196/15411
74. Pini N, Lucchini M, Esposito G, Tagliaferri S, Campanile M et al (2021) A machine learning approach to monitor the emergence of late intrauterine growth restriction. Front Med Sci. https://doi.org/10.3389/frai.2021.622616
75. Crockart IC, Brink LT, Plessis CD, Odendaal HJ (2021) Classification of intrauterine growth restriction at 34–38 weeks gestation with machine learning models. Inform Med Unlocked. https://doi.org/10.1016/j.imu.2021.100533
76. Teng LY, Mattar CNZ, Biswas A, Hoo WL, Saw SN (2022) Interpreting the role of nuchal fold for fetal growth restriction prediction using machine learning. Sci Rep. https://doi.org/10.1038/s41598-022-07883-0
77. Aslam N, Khan IU, Aljishi RF, Alnamer ZM, Alzawad ZM et al (2022) Explainable computational intelligence model for antepartum fetal monitoring to predict the risk of IUGR. Defin Eng Govern Green Artif Intell. https://doi.org/10.3390/electronics11040593
78. Gupta S, Kumar Y (2022) Cancer prognosis using artificial intelligence-based techniques. SN Comput Sci 3(1):1–8. https://doi.org/10.1007/s42979-021-00964-3
79. Carnimeo L (2008) An Intelligent Analyzer for Supporting Diagnoses of Congenital CMV Infection. Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligence, Springer, pp. 1069–1076. https://doi.org/10.1007/978-3-540-85984-0_128
80. Koul A, Bawa RK, Kumar Y (2022) Artificial intelligence in medical image processing for airway diseases. Connected e-Health. Springer, Cham, pp 217–254
81. Boger RA, Boger YS, Foster CB, Boger Z (2008) The use of artificial neural networks in prediction of congenital CMV outcome from sequence data. Bioinform Biol Insights. https://doi.org/10.4137/bbi.s764
82. Tanimura K, Yamada H (2018) Potential biomarkers for predicting congenital cytomegalovirus infection. Int J Mol Sci. https://doi.org/10.3390/ijms19123760
83. Rogers R, Saharia K, Chandorkar A, Weiss ZF, Vieira K, Koo S, Farmakiotis D (2020) Clinical experience with a novel assay measuring cytomegalovirus (CMV)-specific CD4+ and CD8+ T-cell immunity by flow cytometry and intracellular cytokine staining to predict clinically significant CMV events. BMC Infect Dis. https://doi.org/10.1186/s12879-020-4787-4
84. Deepa K, Suganya S (2020) Multiple attribute feature extraction and high support vector classifier for identification of cytomegalovirus images. ICTACT journal on image and video processing. https://doi.org/10.21917/ijivp.2020.0324
85. Hu Z, Tanga A, Singha J, Bhattacharyaa S, Butte AJ (2020) A robust and interpretable end-to-end deep learning model for cytometry data. National Center of Biotechnology Information. https://doi.org/10.1073/pnas.2003026117
86. Lee JS, Yun J, Ham S, Park H, Lee H, Kim J, Byeon JS et al (2021) Machine learning approach for differentiating cytomegalovirus esophagitis from herpes simplex virus esophagitis. Sci Rep. https://doi.org/10.1038/s41598-020-78556-z
87. Eisenberg L, Brossette C, Rauch J, Grandjean A, Ottinger H, Rissland J, Schwarz U, Graf N et al (2022) Time-dependent prediction of mortality and cytomegalovirus reactivation after allogeneic hematopoietic cell transplantation using machine learning. MedRxiv. https://doi.org/10.1101/2021.09.14.21263446
88. Jiang D, Hao M, Ding F, Fu J, Li M (2018) Mapping the transmission risk of Zika virus using machine learning models. Acta Trop. https://doi.org/10.1016/j.actatropica.2018.06.021
89. Mahalakshmi B, Suseendran G (2019) Prediction of zika virus by multilayer perceptron neural network (MLPNN) using cloud. International Journal of Recent Technology and Engineering (IJRTE). https://doi.org/10.35940/ijrte.B1041.0982S1119
90. Lusk R, Zimmerman J, Maldeghem KV, Kim S, Roth NM, Lavinder J et al (2020) Exploratory analysis of machine learning approaches for surveillance of Zika-associated birth defects. Birth Defects Res. https://doi.org/10.1002/bdr2.1767
91. Herry CL, Soares HMF, Faccini LS, Frasch MG (2021) Machine learning model on heart rate variability metrics identifies asymptomatic toddlers exposed to zika virus during pregnancy. Physiol Meas. https://doi.org/10.1088/1361-6579/ac010e
92. Veiga RV, Faccini LS, França GVA et al (2021) Classification algorithm for congenital Zika Syndrome: characterizations, diagnosis and validation. Sci Rep. https://doi.org/10.1038/s41598-021-86361-5
93. Dadheech P, Mehbodniya A, Tiwari S, Kumar S, Singh P, Gupta S (2022) Zika virus prediction using ai-driven technology and hybrid optimization algorithm in healthcare. J Healthcare Eng. https://doi.org/10.1155/2022/2793850
94. Tejera E, Areias MJ, Rodrigues A, Ramõa A, Nieto-villar JM, Rebelo I (2011) Artificial neural network for normal, hypertensive, and preeclamptic pregnancy classification usingmaternal heart rate variability Indexes. J Matern Neonatal Med. https://doi.org/10.3109/14767058.2010.545916
95. Moreira MWL, Rodrigues JLPC, Al-Muhtadi J, Korotaev VV, Albuquerque VHC (2018) Neuro-fuzzy model for HELLP syndrome prediction in mobile cloud computing environments. Concurrency Comput Pract Expert. https://doi.org/10.1002/cpe.4651
96 Tahir M, Badriyah T, Syarif I (2018) Classification algorithms of maternal risk detection for preeclampsia with hypertension during pregnancy using particle swarm optimization. Emitter Int J Eng Technol. https://doi.org/10.24003/emitter.v6i2.287
97. Bennett R, Mulla ZD, Parikh P, Hauspurg A, Razzaghi T (2022) An imbalance-aware deep neural network for early prediction of preeclampsia. PLoS ONE. https://doi.org/10.1371/journal.pone.0266042
98. Kumar Y, Gupta S (2022) Deep transfer learning approaches to predict glaucoma, cataract, choroidal neovascularization, diabetic macular edema, drusen and healthy eyes: an experimental review. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-022-09807-7
99. Kumar Y, Gupta S, Singla R, Hu YC (2022) A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. Arch Comput Methods Eng 29(4):2043–2070. https://doi.org/10.1007/s11831-021-09648-w
100. Kumar Y, Koul A, Mahajan S (2022) A deep learning approaches and fastai text classification to predict 25 medical diseases from medical speech utterances, transcription and intent. Soft Comput 26(17):8253–8272. https://doi.org/10.1007/s00500-022-07261-y
101. Kumar Y, Patel NP, Koul A, Gupta A (2022) Early Prediction of Neonatal Jaundice using Artificial Intelligence Techniques. 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2, 222–226. https://doi.org/10.1109/ICIPTM54933.2022.9753884
102. Kumar Y, Singla R (2022) Effectiveness of machine and deep learning in iot-enabled devices for healthcare system. In: Ghosh U, Chakraborty C, Garg L, Srivastava G (eds) Intelligent internet of things for healthcare and industry. Springer, Cham, pp 1–19
103. Zhong H, Xiao J (2019) Retracted: enhancing health risk prediction with deep learning on big data and revised fusion node paradigm. Hindawi. https://doi.org/10.1155/2019/9757658
104. Uddin S, Khan A, Hossain ME, Moni MA (2019) Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. https://doi.org/10.1186/s12911-019-1004-8

105. Katarya R, Srinivas P (2020) Predicting heart disease at early stages using machine learning: A survey. In 2020 International Conference on electronics and sustainable communication systems (ICESC), pp. 302–305. https://doi.org/10.1109/ICESC48915.2020.9155586
106. Sharma M, Swati, Vig L (2018) Automatic Chromosome Classification using Deep Attention Based Sequence Learning of Chromosome Bands. in 2018 International Joint Conference on Neural Networks (IJCNN), IEEE. https://doi.org/10.1109/IJCNN.2018.8489321
107. Kohli PS, Arora S (2018) Application of machine learning in disease prediction. In 2018 4th International Conference on computing communication and automation (ICCCA), pp. 1–4. https://doi.org/10.1109/CCAA.2018.8777449
108. Yuan FQ (2016) Critical issues of applying machine learning to condition monitoring for failure diagnosis. IEEE International Conference on industrial engineering and engineering management (IEEM). https://doi.org/10.1109/IEEM.2016.7798209
109. Ismaeel S, Miri A, Chourishi D (2015) Using the extreme learning machine (elm) technique for heart disease diagnosis. IEEE Canada International humanitarian technology conference (IHTC2015). https://doi.org/10.1109/IHTC.2015.7238043
110. Dengju Y, Yang J, Zhan X (2013) A novel method for disease prediction: Hybrid of random forest and multivariate adaptive regression splines. Journal of Computers (Finland). https://doi.org/10.4304/jcp.8.1.170-177
111. Dessi A, Ottonello G, Fanos V (2012) Physiopathology of intrauterine growth retardation: from classic data to metabolomics. J Matern Fetal Neonatal Med. https://doi.org/10.3109/14767058.2012.714639
112. Lakshmi BN, Indumathi TS, Ravi N (2015) Prediction based health monitoring in pregnant women. International Conference on Applied and Theoreti- cal Computing and Communication Technology (iCATccT), IEEE. https://doi.org/10.1109/ICATCCT.2015.7456954
113. Zea-Vera A, Ochoa TJ (2015) Challenges in the diagnosis and management of neonatal sepsis. J Trop Pediatr. https://doi.org/10.1093/tropej/fmu079
114. Hou B, Khanal B, Alansary A, McDonagh S, Davidson A, Rutherford M, Hajnal JV et al (2018) 3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2-D images. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2018.2798801
115. He F, Lin B, Mou K, Jin L, Liu J (2021) A machine learning model for the prediction of down syndrome in second trimester antenatal screening. Clin Chim Acta. https://doi.org/10.1016/j.cca.2021.07.015
116. Jamshidnezhad A, Hosseini SM, Mohammadi-Asl J, Mahmudi M (2021) An intelligent prenatal screening system for the prediction of Trisomy-21. Inform Med Unlocked. https://doi.org/10.1016/j.imu.2021.100625
117. Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P (2021) Prediction of heart disease using a combination of machine learning and deep learning. Comput Intell Neurosci. https://doi.org/10.1155/2021/8387680
118. Rani S, Masood S (2020) Predicting congenital heart disease using machine learning techniques. J Discret Math Sci Cryptogr. https://doi.org/10.1080/09720529.2020.1721862
119. Clifford GD, Silva I, Behar J, Moody GB (2017) Non-invasive fetal ECG analysis. Physiol Meas. https://doi.org/10.1088/0967-3334/35/8/1521
120. Bauer ME, Bateman BT, Bauer ST, Shanks AM, Mhyre JM (2013) Maternal sepsis mortality and morbidity during hospitalization for delivery: temporal trends and independent associations for severe sepsis. Anesth Analg. https://doi.org/10.1213/ANE.0b013e3182a009c3
121. Bansal K, Bathla RK, Kumar Y (2022) Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer. Soft Comput 26(21):11153–11184. https://doi.org/10.1007/s00500-022-07246-x
122. Khalilia M, Chakraborty S, Popescu M (2011) Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak. https://doi.org/10.1186/1472-6947-11-51
123. Vinitha S, Sweetlin S, Vinusha H, Sajini S (2018) Disease prediction using machine learning over big data. Int J Comput Sci Eng. https://doi.org/10.2139/ssrn.3458775
124. Patil M, Lobo VB, Puranik P, Pawaskar A, Pai A, Mishra R (2018) A proposed model for lifestyle disease prediction using support vector machine. In 2018 9th International Conference on computing, communication and networking technologies (ICCCNT). https://doi.org/10.1109/ICCCNT.2018.8493897