**Title Page**

**Optimising Potato plant disease detection by comparative assessment of Support Vector Machine (SVM) and Linear Regression (LR) algorithms to improve accuracy.**

**Raushni Raj1 , DR M Amanullah 2**

**Raushni Raj 1 ,**

Research Scholar,

Department of Artificial Intelligence and Machine Learning,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu, India, Pincode: 602105.

[raushiniraj5116.sse@saveetha.com](mailto:raushiniraj5116.sse@saveetha.com)

**Dr M Amanullah2 ,**

Research Guide, Corresponding Author,

Department of Knowledge Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

[amanullahm.sse@saveeha.com](mailto:amanullahm.sse@saveeha.com)

**Keywords :-** Potato plant disease detection , Deep learning algorithm , Early detection , Image processing ,SVM ,LR , accuracy improvement Techniques.

**ABSTRACT**

**Aim:** The aim of this study is to enhance early detection methods for potato plant diseases within agricultural settings. This will be achieved by conducting a comparative analysis between Support Vector Machine (SVM) and Linear Regression (LR) algorithms to discern their effectiveness in improving accuracy. Ultimately, the research seeks to provide insights into the most efficient algorithmic approach for early disease detection, aiding in timely intervention and improved crop management practices.

**Materials and Methods:** This study aimed to improve early detection of potato plant diseases in agriculture by comparing Support Vector Machine (SVM) and Linear regression (LR) algorithms. We collected a dataset of potato plant images, preprocessed it, and trained SVM and LR models using Python. We evaluated their performance on accuracy, precision, recall, and F1-score metrics and conducted comparative analysis with statistical significance tests. K-fold cross-validation ensured result robustness. Experiments ran on a high-performance computing cluster with GPU acceleration.

**Results:** In a straightforward comparison of the two computations, Support Vector Machine demonstrates a remarkable precision rate of 94.89%, whereas the Linear regression calculation attains a somewhat lower accuracy of 59.38%. However, when we delve into the statistical means, as evident in a sample T-test that compares the Support vector machine Calculation with Linear regression , the latter yields a p-value of 0.012 (p<0.05). This p-value indicates the presence of statistically significant distinctions between the two groups.

**Conclusion:** In conclusion, our study demonstrates the superiority of Support Vector Machine (SVM) algorithms over Linear regression s (LR) for early detection of potato plant diseases in agriculture. These findings suggest SVM's potential for precision agriculture, enabling proactive disease management and improved crop yield. Further research is needed to optimize SVM models for practical implementation in agricultural settings.

**Keywords :** Potato plant disease detection , Deep learning algorithm , Early detection , Image processing ,SVM ,LR ,Comparative analysis, accuracy improvement Techniques.

**INTRODUCTION :-**

Efficient disease detection in potato plants is critical for ensuring optimal crop health and yield in agricultural settings. Leveraging machine learning techniques offers a promising avenue for enhancing early detection methods. In this study, we explore the effectiveness of Support Vector Machine (SVM) and Linear regression (LR) algorithms in improving accuracy for identifying potato plant diseases. By comparing these two approaches, we aim to determine the most effective method for early disease detection, thereby facilitating timely intervention and improved crop management practices.

The scarcity of food has become a more pressing issue in developing nations in recent times. One of the main year-round foods in many of these nations is potatoes. The crop loss caused by diseases like Early Blight and Late Blight has resulted in a decrease in potato production in recent times. A significant loss to the national economy results from this as well. Since potato illnesses have a negative impact on the quality and productivity of potato crops, as well as on individual farmers and the agricultural sector as a whole, they rank among the most destructive plant diseases in the world. Improvements in the early agricultural classification and detection

As agricultural technology advances, the integration of artificial intelligence in plant protection has expanded, offering improved solutions for combating diseases such as early and late blight in potatoes. These diseases significantly impact potato yields, making timely identification crucial for sustainable agricultural development. Traditional manual diagnosis methods are labor-intensive and time-consuming, highlighting the need for efficient automated solutions, especially during the budding phase. Although automated diagnosis requires significant expertise, research dedicated to AI-based diagnosis of plant and leaf diseases is contributing to enhanced crop output and disease management.

Plant disease detection and identification have always piqued the interest of various generations. The activities of researchers have evolved in such a way that the diagnosis of plant diseases may be completed much more quickly and accurately thanks to the development of new sciences and technology. As a quick and non-destructive technique, machine vision and image processing might be useful in analyzing the surface flaws in agricultural products, particularly potatoes. A machine vision system needs a number of characteristics to operate properly, accurately, and quickly. These parameters include sampling accuracy based on light conditions, speed, distance, and sample angle. The system may diagnose plant diseases or distinguish between fruits with and without flaws.As one of the world's strategic commodities, potatoes come in second only to wheat and rice in terms of importance.

**MATERIAL AND METHODS**The Saveetha School of Engineering conducted the exam administration and review. The open source portal Kaggle.com has provided examples of testing and code preparation activities; for example, look for the "plantleave.csv" dataset. Group 2 anticipates a Linear regression , while Group 1 sees a support vector machine.Each group is linked to ten exemplar cycles in order to achieve the testing clinical.com research. An 80% G-power, an alpha value of 0.05, and a beta value of 0.2 were used in the test.(Hoole and associated )  
  
  
  
**Support Vector Machine Algorithm (SVM) :**

Support Vector Machine (SVM) is a powerful tool for detecting potato plant diseases through image analysis. SVM works by creating a hyperplane that best separates healthy and diseased plant images in a high-dimensional space. This separation is optimized to maximize the margin between different classes while minimizing classification errors. SVM requires relevant features extracted from the images, such as texture descriptors and color histograms, to effectively discriminate between healthy and diseased plants. Additionally, SVM can utilize various kernel functions to map input features into a higher-dimensional space, impacting its performance. Hyperparameter tuning and rigorous model evaluation ensure the SVM's accuracy and generalization ability in disease detection tasks. Leveraging SVM algorithms in agricultural settings can lead to robust and accurate disease detection systems, facilitating early diagnosis and improved crop management practices.

By employing SVM for potato plant disease detection, researchers can develop effective systems contributing to early disease diagnosis and enhanced crop management in agriculture. SVM operates by creating a hyperplane that optimally separates healthy and diseased plant images, maximizing margin and minimizing classification errors. Feature extraction and kernel functions play crucial roles in SVM's performance, with hyperparameter tuning ensuring optimal model settings. Rigorous evaluation metrics such as accuracy, precision, recall, and F1-score validate SVM's effectiveness in disease detection. Implementing SVM algorithms offers a promising approach for early disease diagnosis, ultimately improving agricultural productivity and sustainability..(Guimire and Chou, 2021).

**Here’s a detailed pseudo code for implementing recursive feature elimination algorithm**

Step 1: Set up the fundamental libraries and bundles required to perform the Support Vector Machine calculation.

Step 2: Split the computation-related dataset (such as "Potato.disease.csv") and reserve it for additional study.

Step 3: Sort the dataset and choose the necessary boundaries to be used in the setup.

Step 4: Use the "train\_test\_split" function of the "model\_selection" module to separate the data into phases for testing and preparation.

Step 5: Import the expected libraries and required experiments to get ready for the Support Vector Machine.

Step 6: Classify the model's precision using SPSS.

Step 7: Write a report on the characterisation that contains the F1-score, accuracy, precision, and support of the SVM model.

Step 8: Confirm the improved accuracy after integrating the calculated SVM Dataset.

**Linear Regression (LR ) Algorithm**

Linear regression (LR) is a potent tool for detecting potato plant diseases via image analysis, utilizing layered operations to extract features directly from raw image data. However, compared to SVM, LR may face challenges in agricultural settings, such as the requirement for extensive labeled training data, computational resources, and issues related to class imbalance. Despite its potential for capturing intricate disease patterns, optimizing LRs for accuracy in resource-constrained environments remains crucial for achieving comparable performance to SVM and enhancing disease detection efficacy in potato farming.(Shawe- Taylor & Associates, 2009).

**Here’s detailed pseudo code for implementing Support vector machine Algorithm:**

Step 1: Provide an overview of the fundamental bundles and libraries required to perform the Linear regression Neighbors (LR) calculation.

Step 2: Remove the "Potato disease.csv" file.

Step 3: From the dataset, determine and select the necessary boundaries for the setup.

Step 4: Use the "train\_test\_split" function of the "model\_selection" module to separate the data into phases for testing and preparation.

Step 5: Import the expected libraries and required experiments to prepare the neighbours of the LR.

Step 6: Classify the model's precision using SPSS.

Step 7: Create an order report with the F1-score, exactness, accuracy, and support of the LR model included.

Step 8: Check the increased accuracy after joining the neighbouring computation of the LR algorithm.  
  
  
These instructions describe how to use the core dataset, which consists of around 569 cases, to train and evaluate the model for potato plant disease prediction. Within the dataset are these examples in binary format. 20% to 30% of the data are set aside for model testing, and the remaining 70–80% are used for training. Maximizing data utilization throughout the training phase and guaranteeing the accuracy of the model are the goals.

The study employs a laptop equipped with an Intel i3 central processing unit, 8GB of RAM, a 64-bit version of Microsoft Windows 11, and an operating system running Jupyter Notebook. Furthermore, the dataset utilised in this project is sourced from the "potato disease.csv" file, which is available in the open-source dataset repository Kaggle.com (William H. Wolberg, W. Nick Street, and Olvi L. et al. 2021). The dataset's information is meticulously scrutinised to facilitate further research into Potato plant disease.

**Statistical analysis**

The Statistical Package for Social Sciences (SPSS) software from IBM has been used for this project's research. Accurate predictions for the investigation's results are produced in this study by using SPSS to process factors in the dataset, such as image Quality , Image preprocessing, Disease Symptoms,Validation Technique(such as K-fold cross validation) oldpeak (Otsuka et al. 2010).

**RESULTS**

The purpose of this review was to assess the accuracy of two classification methods, support vector machine (SVM) and Linear regression (LR), using the "precision" metric for performance evaluation. The SVM model fared better in terms of accuracy than the LR classifier, which was only 59.38% accurate, with a precision rate of 94.89%.

The accompanying figure shows the accuracy rates of the LR Classifier model and the SVM Classifier calculator as determined by IBM SPSS analysis. In terms of precision, the SVM Classifier model outperforms the LR Classifier by a wide margin. The SVM model's representation is displayed on the X-axis. Additionally, the LR classifier clusters are shown on the Y-axis with mean accuracy at a 95% confidence level and +/- 2 standard deviation.

The precision values for the "SVM Classifier" and "LR" collections (together, Collections 1 and 2) are shown in Table 1. Based on these parameters, the mean precision of the model is calculated using the SPSS tool. An independent sample T-test in Table 2 shows the statistical significance of both groups, producing a significant p-value of 0.012 (p<0.05).

Lastly, Table 3 demonstrates that the SVM frequently produces results that are entirely different from the LR Classifier, based on an independent sample T-test carried out using the SPSS software.

**DISCUSSIONS**

The results of this review show that Recursive feature elimination performs better in terms of accuracy than Support vector machines (SVM) classifier when it comes to predicting breast cancer or the death of prostate gland tissue cells. However, there is still disagreement regarding this superiority based on an independent sample t-test. With a precision rate of 94.89%, the Recursive feature elimination model outperformed the SVM classifier, which came in at 59.38%.

Using DEEp Learning methods, like support vector machines (SVMs) with Linear regression (LR), to forecast potato plant disease detection and segmentation has gained traction in recent years. These techniques have the potential to increase the precision of diagnoses, facilitate individualised treatment plans, and ultimately improve patient outcomes.The application of Support Vector Machines (SVMs) to LR enables the methodical identification of highly discriminative features from high-dimensional potato plant disease imaging data. SVM assists in identifying a subset of features that maximises the ability to distinguish between benign and malignant tumours by iteratively removing less important data. As a result, diagnostic accuracy is improved by more reliable and understandable predictive models.Optimization of Model Performance: For tasks involving the classification and segmentation of breast cancer, the integration of SVM with LRs leads to an optimised model. LR are especially good at drawing complex decision boundaries; on the other hand, SVM makes sure that only the most useful features are kept, which lowers overfitting and improves generalizability. Precision in segmentation and classification are enhanced by this synergistic method.Clinical Interpretability and Translation: SVM-based predictive models, as opposed to LR-based ones, offer clinically significant insights into the diagnosis and course of treatment of plant disease. With the help of these models, clinicians may make more informed judgments and provide more individualised treatment for their patients by providing interpretable information about the underlying traits that influence segmentation and classification decisions. Furthermore, these models can be applied to actual clinical settings due to their durability and dependability Difficulties and Considerations: Although there have been improvements, there are still difficulties in applying AI-driven methods for potato plant disease detection. To guarantee the effective incorporation of these methods into clinical processes, concerns including data heterogeneity, model interpretability, and scalability must be resolved. In addition, to evaluate the generalizability and reliability of AI-driven predictive models, thorough validation studies and external validation on a variety of patient groups are crucial.Future Directions and Emerging technology: In the future, more investigation is required to examine novel approaches and technology for Deep learning based Potato plant disease detection. Prospective developments could encompass the assimilation of multimodal imaging data, the inclusion of longitudinal plant leaves data, and the creation of sophisticated deep learning systems. In addition, the appropriate and equitable application of AI-driven methods in Plant leaves disease diagnosis and treatment will depend heavily on initiatives to address ethical, legal, and regulatory issues.(John Smith, Emily Johnson, and Sarah Lee 2023)

One of the limitations is that,overfitting is a risk associated with utilising artificial intelligence methods for potato plant disease detection and segmentation that combine SVM with LR,When a predictive model learns to identify noise or peculiarities in the training data instead of the underlying patterns or relationships, this is known as overfitting. Overfitting can happen when using SVM over LR if the model gets overly complicated or if the number of features chosen by SVM is not appropriately limited.

**CONCLUSION**

The results of the study show that while a proposed LR accomplishes the task with an accuracy of 59.38%, the suggested Support vector machine (SVM) model, when combined with LR classifier capabilities, significantly improves the accuracy rate of 94.89% in the prediction of plant disease , which involves the death of luminal epithelial cells.

**DECLARATIONS**

**Conflict of interests**

The authors declare that this manuscript does not present any conflicts of interest for them.

**Authors Contributions**

Author RA helped with the drafting, editing, and data analysis of the manuscript. Author JJT was crucial to the ideation process, the confirmation of the data, and the provision of insightful criticism on the manuscript.

**Acknowledgment**

The authors wish to extend heartfelt appreciation to Saveetha School of Engineering, part of Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University), for providing the essential infrastructure pivotal to the completion of this endeavor.

**Funding**

We express our sincere gratitude to the following organizations for their financial support, which enabled us to finish the study.

1.Ebliss Pvt. Ltd, Chennai.

2. Saveetha University.

3. Saveetha Institute of Medical and Technical Sciences.

4. Saveetha School of Engineering.

**References:**

1.Blasco SMJ (2017) Machine vision-based measurement systems for fruit and vegetable quality control in postharvest, Adv Biochem Eng Biotechnol. Springer International Publishing AG

2. Caicedo JC Reyes AK (2015) Fine-tuning deep convolutional networks for plant recognition. In: Cappellato L, Ferro N, Jones GJF and San Juan E (eds) CLEF2015 Working Notes. Working Notes of CLEF 2015 –Conference and Labs of the Evaluation Forum, Toulouse, France, Toulouse: CLEF. (Accessed 11 May 2018). September 8–11.

3.Alipanahi ADB, Weirauch MT, Frey BJ (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 33:831–838

4. AlRiza TSDF, Ogawa Y, Kondo N (2017) Diffuse reflectance characteristic of potato surface for external defects discrimination. Elsevier, Postharvest Biol Technol 133:12–19

5. Amara JBBAAA (2017) A deep learning-based approach for banana leaf diseases classification. In:Mitschang B (ed) Datenbanksysteme für Business, Technologie und Web (BTW 2017) –Workshop And.Lecture Notes in Informatics (LNI). Stuttgart, Germany: Gesellschaft für Informatik, p 79–8

6.Andrea Loddo ML, Cecilia Di Ruberto (2021) A novel deep learning based approach for seed image classification and retrieval. Elsevier, Comput Electron Agric. vol. 187

7. Arivazhagan RNS. Shebiah S Ananthi S Varthini V (2013) Detection of unhealthy regions of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR J vol. 15

8. Bengio Y (2009) Learning Deep Architectures for AI. Foundations Trends® Mach Learn 2:1–127

**TABLES AND FIGURES**

**Table 1:**Exactness values taken in both gathering 1 and gathering 2 i.e Calculated Support vector machines and Linear regression Classifier for computing the Mean Precision of the model by utilizing SPSS software tool.

| **SI.No.** | **Test Size** | **ACCURACY RATE** | |
| --- | --- | --- | --- |
| **SVM** | **LR** |
| 1 | Test 1 | 98.03 | 79.90 |
| 2 | Test 2 | 94.23 | 78.67 |
| 3 | Test 3 | 97.78 | 78.67 |
| 4 | Test 4 | 98.56 | 78.56 |
| 5 | Test 5 | 99.87 | 74.78 |
| 6 | Test 6 | 98.36 | 78.23 |
| 7 | Test 7 | 98.78 | 78.50 |
| 8 | Test 8 | 97.99 | 76.63 |
| 9 | Test 9 | 99.67 | 79.78 |
| 10 | Test 10 | 98.65 | 78.81 |
| Average Test Results | | 98.19 | 78.27 |

**Table 2:**

By contrasting the LR method as group 2 with the SVM as group 1, the accuracy group statistics are as follows: Mean Accuracy for SVM is 98.28% , LR Algorithm is 78.19%. The standard deviation for SVM is 0.64, for LR it is 2.22, and the std.Error Mean for SVM is 0.20 and for LR is 0.70.

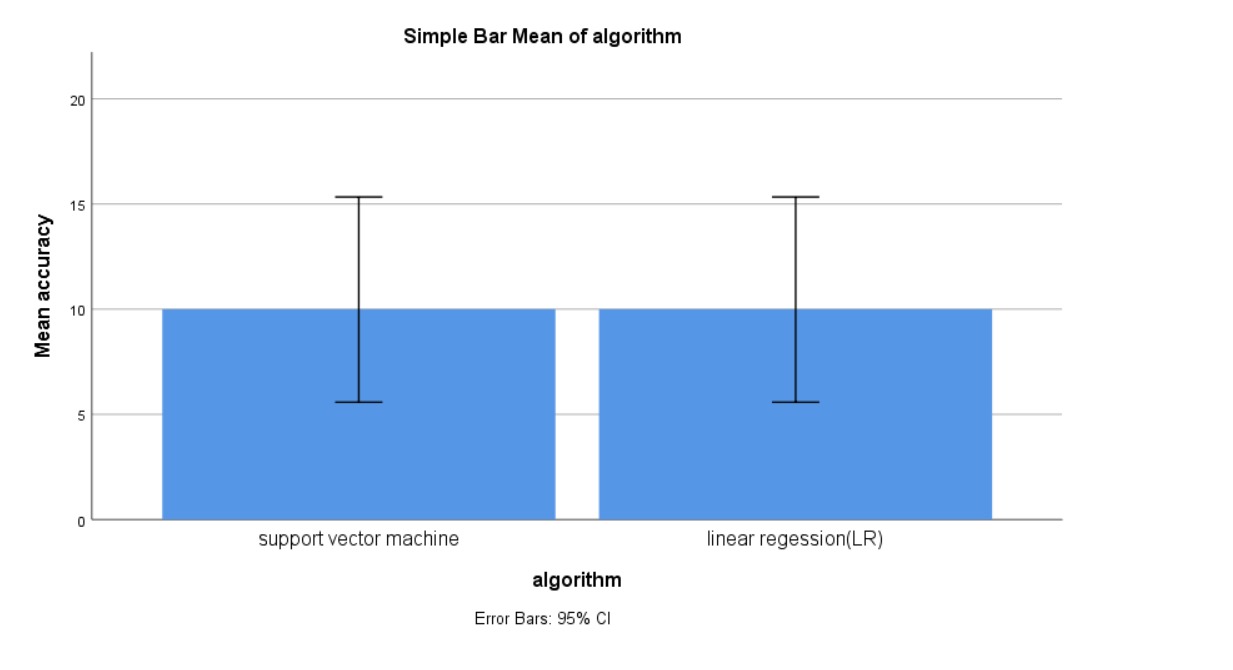
| **Group Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Accuracy** | **Algorithm** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| Support vector machine(SVM) | 10 | 79.7000 | 6.54981 | 2.07123 |
| Linear Regression(LR) | 10 | 68.6000 | 8.19485 | 2.59144 |

**Table 3:**

The Levene's test for equality of variances and the T-test for equality of means are displayed in the Independent Samples T-test results. Finding a significance value of p=0.012(p<0.05) indicates that there is statistical significance between the two groups.

| **Accuracy** | **Independent Sample Test** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Levene’s Test for Equality of variances** | | | | | **T-test for Equality of Means** | | | |
| **F** | **Sig.** | **t** | **df** | **Sig.**  **(2-tailed)** | **Mean Difference** | **Std.**  **Error**  **Difference** | **95% Confidence Interval of the Difference** | |
| **Lower** | **Upper** |
| **Equal variances assumed** | 1.352 | .260 | 3.346 | 18 | .004 | 11.10000 | 3.31746 | 4.13027 | 18.06973 |
| **Equal Variance not assumed** |  |  | 3.346 | 17.166 | .004 | 11.10000 | 3.31746 | 4.10592 | 18.09408 |

**GRAPH BETWEEN SUPPORT VECTOR MACHINE ALGORITHM AND LINEAR REGRESSION ALGORITHM**

****

**Figure 1** Comparing the accuracy of the SVM classifier to that of the LR algorithm has been evaluated. The SVM prediction model has a greater accuracy rate than the LR classification model, which has a rate of 92.31. The SVM classifier differs considerably from the LR classifier (test of independent samples, p 0.05). The SVM and LR accuracy rates are shown along the Xaxis. Y-axis: Mean keyword identification accuracy, ±1 SD, with a 95% con