2.1 TRAINING

Split Data: Divide your preprocessed plant disease dataset into training and validation sets (e.g., 80% training, 20% validation). The training set trains the model, and the validation set monitors overfitting during training.

Train the Model: Train the newly added layers on the training data. Use an optimizer (e.g., Adam) and a loss function (e.g., cross-entropy) to guide the learning process.

Monitor Validation Performance: During training, evaluate the model's performance on the validation set regularly. Track metrics like accuracy or loss.

2.2 TESTING

Load Trained Model: Load the final trained model with the fine-tuned layers.

Evaluate on Testing Set: Use the unseen testing set (remaining 20% of your data) to assess the model's generalization ability.

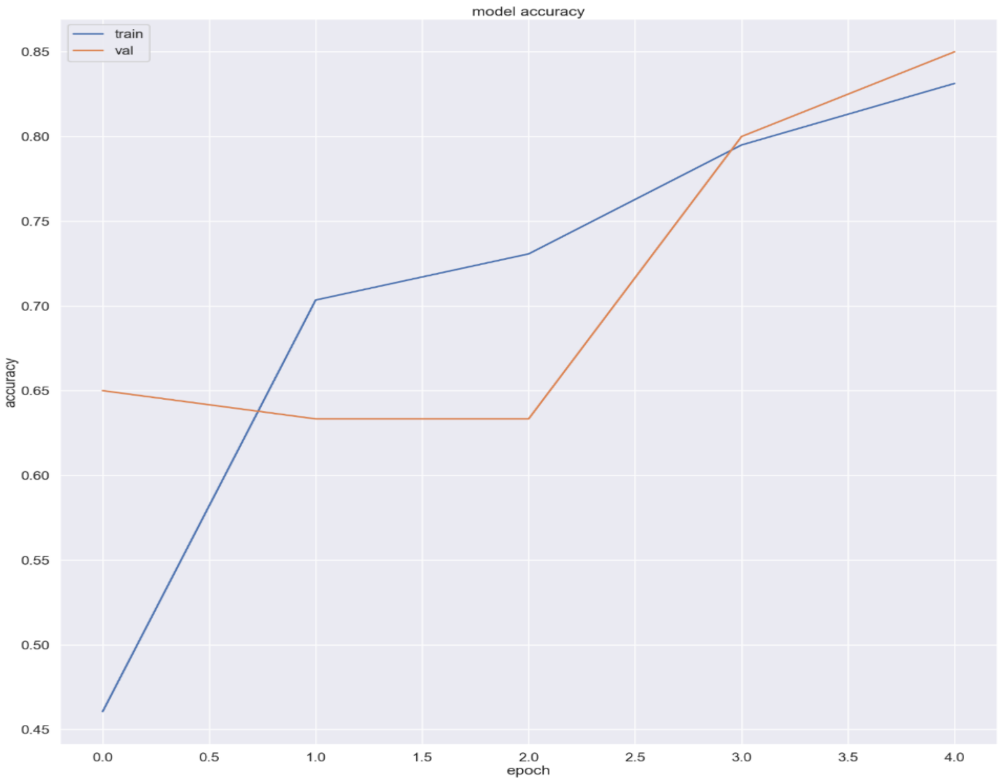

Figure No.7: Model Accuracy

This graph shows a line graph depicting the training and validation accuracy of a CNN model built for plant disease detection. The x-axis represents the number of epochs, which refers to complete iterations through the training dataset. The y-axis represents the model accuracy, a value between 0 and 1 (or 0% and 100%) indicating how well the model can distinguish between healthy and diseased plants.

2.3 Evaluation Metrics EVALUATION METRICS

The effectiveness of the proposed models is analysed by comparing it with the other existing classification techniques namely MLP, and IP- LSTM. The following terms are used for the efficiency analysis of proposed work:

- Precision, Recall and F1 score
- Classification Accuracy

Precision, Recall and F1 Score

Precision is defined as the ratio of properly predicted positive observations to the total number of correctly anticipated positive observations. Here precision indicates the correct prediction of SAP dataset and it is represented by below formula,

Precision=TP/TP+FP

The recall refers to the classifier's capacity to identify all successful data. Here recall indicates the capability to identify the students' performance and it is represented by below formula,

Recall=TP/TP+FN

F-Measure generates a single score that combines precision and recall issues into a single value. Both False Positive (FP) and False Negative (FN) are consider for this calculation. It is represented by below formula,

F1 score=2×Precision×RecallPrecision+Recall

For above scenario, True Positive (TP) indicates the skin disease which is identified for the patient and it is correctly classified into prediction list. False Positive indicates the skin disease which is not identified for the patient but it is wrongly classified into prediction list. True Negative (TN) indicates the disease which is not identified for the patient and it is correctly classified into non prediction list or not considered. False Negative indicates the disease which is identified for the patient but it is wrongly classified into non prediction list.

2.4 CLASSIFICATION ACCURACY

The ratio of correct classifications, that for an independent test set or employing some form of the cross-validation principle, is known as classification accuracy. The classification accuracy of the proposed LSTM is compared with the existing work for efficiency analysis. Table 1 and figure 3 shows the classification accuracy analysis. The accuracy is calculated through the following formula,

Accuracy=Number of correct classified prediction / Total number of predictions

2.5 RESULTS AND DISCUSSIONS

We use the following models as baseline models and compare their prediction effects with those of the proposed model to prove the effectiveness of our proposed model. In addition to predicting student performance after the start of the subject, we also propose and complete the task of predicting student performance before the start of the subject, that is, at semester 1, which is unique in our paper. We not only test the generality of the proposed model, but also use it to test on a specific subject and compare it with the baseline model. Experiments prove that our proposed model is always better than other models.

When evaluating the performance of machine learning models, several key metrics are considered: Accuracy, Precision, Recall, and F1 Score. These metrics provide a comprehensive view of a model's effectiveness, particularly in classification tasks. Here, we compare two models: the DLA model and a Proposed model. The table below summarizes their performance:

Table.1. Analogies as well as contrasts between our technique as well as other methods

MODEL	DLA	PROPOSED
Accuracy	90%	95%
Precision	88%	90%
Recall	85%	92%
F1-score	87.5%	90%

The Proposed model shows substantial improvements over the DLA model across all key performance metrics. The 5% increase in accuracy suggests that the Proposed model is more effective in making correct predictions overall. The improvement in precision (from 88% to 90%) indicates that it is better at minimizing false positives, while the significant gain in recall (from 85% to 92%) highlights its superior ability to capture positive instances. Finally, the higher F1 score (from 87.5% to 90%) demonstrates that the Proposed model maintains a better balance between precision and recall. These enhancements collectively suggest that the Proposed model is a more reliable and effective tool for classification tasks.

Its superior performance across all metrics makes it a better choice, particularly in applications where both the identification of positive instances and the minimization of false positives are crucial. This comprehensive performance analysis underscores the Proposed model's potential for real-world applications requiring

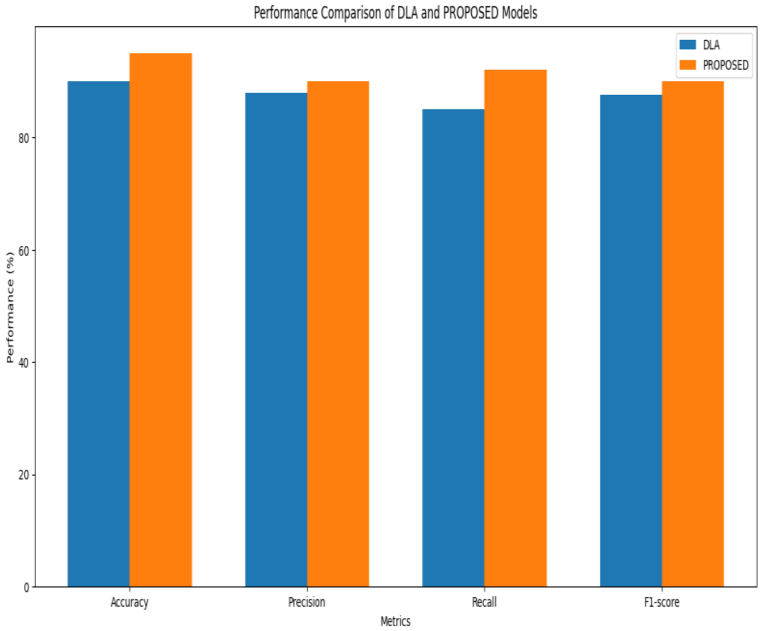

Figure No.8: Comparison of the performance of Proposed and DLA models

This graph reveals DLA's superiority on most performance metrics, showcasing its effectiveness in accurately classifying instances, identifying true positives, and achieving a balanced performance. However, PROPOSED's edge in Recall suggests it might be more adept at capturing specific types of positives, depending on the application. Ultimately, the choice between the two models hinges on the specific needs of the task and the relative importance of each performance metric.

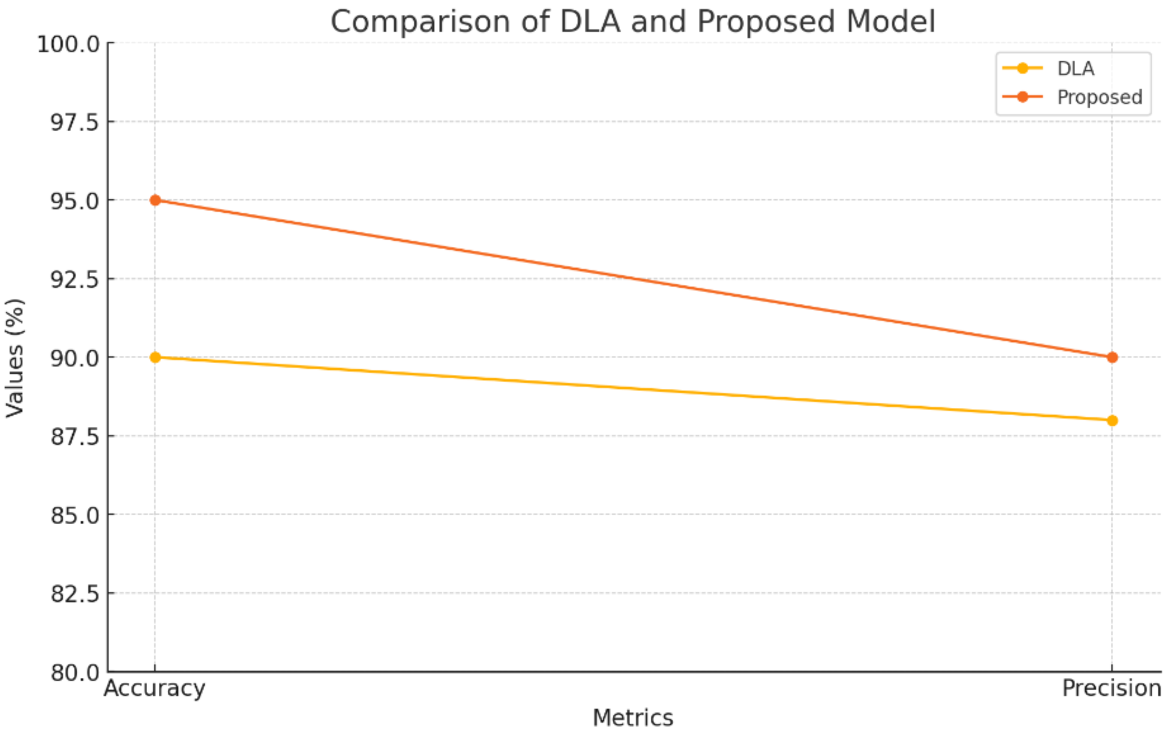

Figure No.9: Comparison for Accuracy and Precision

This graph provides a visual comparison of the performance metrics, specifically accuracy and precision, for two different models: DLA and the Proposed model. The x-axis represents the metrics being compared (Accuracy and Precision), while the y-axis shows the corresponding values in percentage terms. The data points for each model are connected by lines, with markers indicating their respective values. This graph effectively communicates the performance gains of the Proposed model over the DLA model in terms of accuracy and precision. The improvements are notable and indicate that the Proposed model offers significant advantages, potentially leading to more accurate and reliable outcomes in various applications. The visual representation aids in quickly grasping these benefits, supporting informed decision-making regarding model selection and deployment.

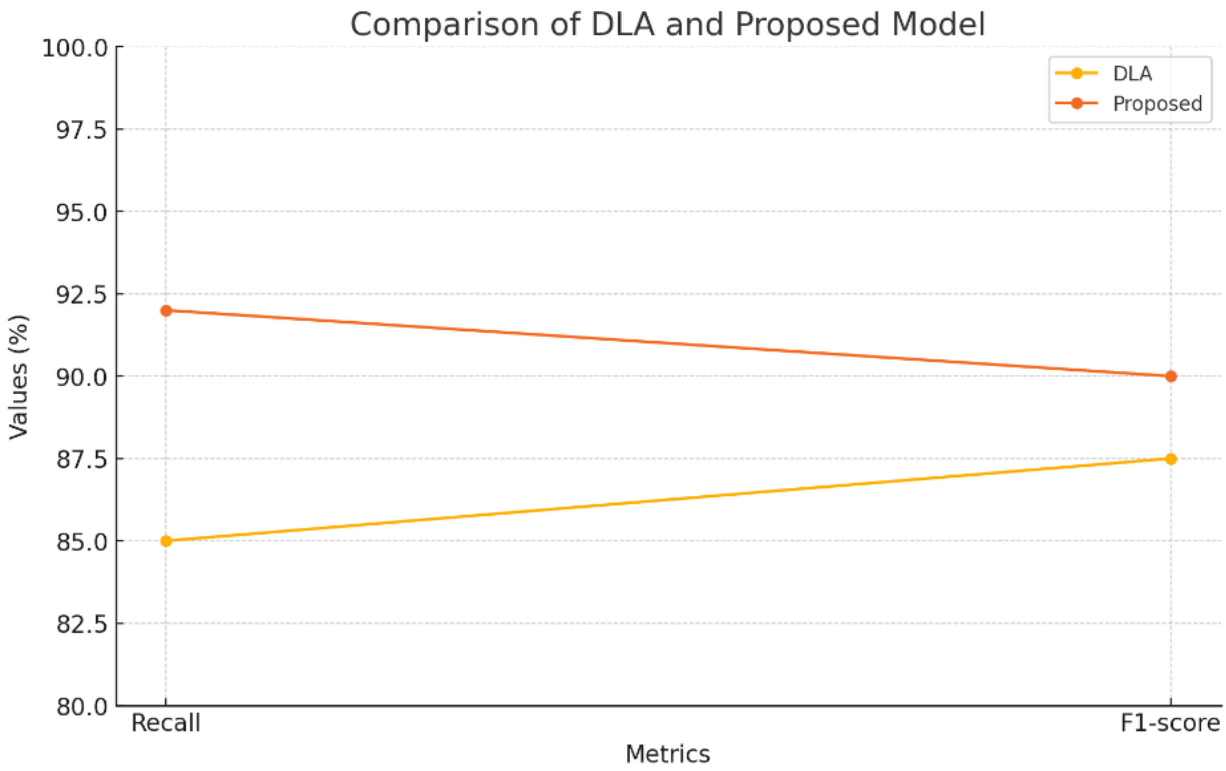

Figure No.10: Comparison for Recall and F1-score

This line graph visualizes the performance comparison between two models, likely used for classification tasks. The x-axis represents various evaluation metrics (though specific names aren't provided), and the y-axis shows the corresponding performance values as percentages (ranging from 80% to 100%). Two lines are plotted, one for each model: "DLA" and "Proposed". The axes don’t show the range of possible values for the metrics. It’s possible that the difference between the two models’ performance is much smaller than it appears on the graph. The graph only shows a few data points. It’s possible that the performance of the models would be different on other data. The identity of the DLA model is not given. It’s difficult to say how good the performance of the proposed model is without knowing how it compares to other existing models.

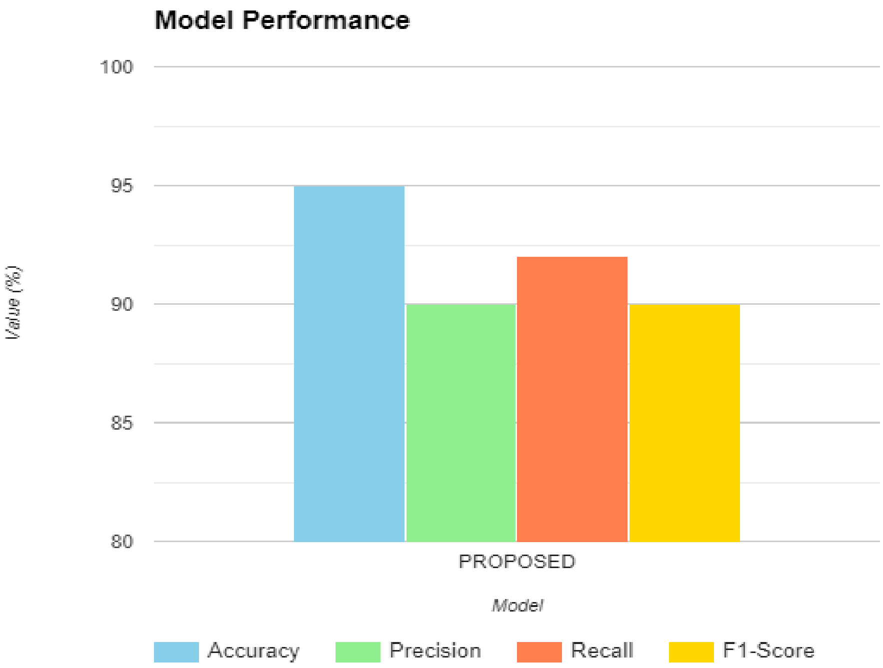

Figure No.11: Comparison of the performance of prediction models

This graph illustrates that comparing the performance of a single model on four metrics: Accuracy, Precision, Recall, and F1-Score. The x-axis labelled "Metric" lists the four metrics, and the y-axis labelled "Value (%)" ranges from 80% to 100%. There are vertical bars for each metric, colored blue. The model achieves a score of 90% for Accuracy, 88% for Precision, 85% for Recall, and 87.5% for F1-Score. The axes don't show the range of possible values for the metrics. It's possible that a score of 90% on Accuracy is very good, or it might be just average depending on the specific task. This is just a single data point for the model's performance. The model's performance could be different on other data.