SHIELD: Symptom-Based Hybrid Intelligent Early Learning for Disease Prediction

'Asrul 'Azeem Bin Fazil Akashah[1], Taniza Binti Tajuddin[2]

[1] *Universiti Teknologi MARA, Malaysia*
[2] *Universiti Teknologi MARA, Malaysia*

Article Information

Received: 21-11-2024
Revised: 28-11-2024
Published: 05-12-2024

Keywords
Hybrid Model; Disease Prediction; Machine Learning; Ensemble Learning; Healthcare

Correspondence Email:
asrlazeem23@gmail.com

Abstract

Traditional diagnostic approaches often face delays and inaccuracies, while standalone machine learning models fail to account for individual uniqueness. The SHIELD system leverages hybrid machine-learning models to enhance disease prediction based on patient symptoms. This study integrates Gradient Boosting, Decision Trees, and Random Forest models, combining their strengths using an ensemble voting approach. A comprehensive dataset from Kaggle, enriched with symptom severity mappings, enables accurate and personalized predictions. The system delivers practical outputs, including disease names, descriptions, and home remedies, through a user-friendly web interface. Achieving an accuracy of approximately 99.59% with the ensemble model, SHIELD demonstrates its potential to revolutionize early disease detection, aligning with global health objectives.

1. Introduction

The global healthcare landscape is witnessing an alarming rise in the prevalence of chronic and infectious diseases, with delayed diagnoses significantly impacting patient outcomes. According to the World Health Organization (WHO), timely disease detection is critical for improving public health and optimizing healthcare resources. However, traditional diagnostic methods are often reactive, costly, and constrained by the limitations of human expertise. Advancements in machine learning (ML) offer an opportunity to revolutionize disease prediction by leveraging large datasets to identify patterns and trends that would otherwise remain undetected. This research explores the potential of integrating hybrid machine-learning models to create an intelligent system for accurate, symptom-based disease prediction.

1.1 Problem Statement

Diagnosing diseases accurately and efficiently remains a challenge in healthcare. Conventional methods often face delays, lack personalization, and rely on models with moderate accuracy. Many existing machine learning models take a "one-size-fits-all" approach, overlooking important patient-specific details like medical history and the severity of symptoms. This results in predictive accuracies that typically range between 80–90%, which is often not reliable enough for real-world clinical use.

To address these issues, this study focuses on answering a few important questions. What symptoms play the biggest role in making accurate predictions? How can we design a machine learning model that takes individual symptoms into account to predict diseases more effectively? And how can combining models—like Gradient Boosting, Decision Trees, and Random Forests—through an ensemble voting approach make predictions more accurate and reliable? By exploring these questions, this research aims to bridge the gap in current diagnostic methods and create a system that is both dependable and personalized for better healthcare outcomes.

1.2 Research Objectives

The growing complexity of diseases and the limitations of conventional diagnostic methods highlight the need for advanced tools that can provide accurate and personalized predictions. Machine learning offers a promising solution, but many existing models fail to account for individual factors, leading to suboptimal outcomes. To address these challenges, this research is driven by the following objectives:

1. To identify the key symptoms that contribute to the prediction of various diseases.
2. To design and develop a machine learning model capable of predicting diseases based on patient symptoms.
3. To evaluate and implement a hybrid ensemble model to improve prediction accuracy and reliability.

These objectives aim to bridge the gaps in existing diagnostic approaches by combining the strengths of advanced machine learning techniques with a focus on personalized healthcare.

1.3 Literature Review

The advancement of data analysis in healthcare has significantly enhanced disease prediction and early intervention. Machine learning (ML) plays a pivotal role in analyzing large datasets, extracting patterns, and enabling decision-making. Traditional data analysis methods often struggle with unstructured data and fail to adapt to the complexity of symptoms and diseases. However, ML offers techniques for processing structured and unstructured data, including missing values and noise, to uncover meaningful insights. This study leverages datasets enriched with symptom severity mappings to prioritize critical symptoms, improving the reliability of disease predictions. Evaluating ML models based on metrics such as accuracy, F1-score, and confusion matrices ensure robust model selection.

2. Approach Method

In the field of disease prediction, various machine-learning methods have been developed to improve accuracy and reliability. These methods can be broadly categorized into predictive models, adaptive models, and hybrid predictive models. Each approach leverages unique strengths to address the limitations of traditional diagnostic systems. By analyzing the performance of these models through comparative studies, their applicability and effectiveness in handling diverse datasets and medical conditions can be better understood.

2.1 Predictive Models

Predictive models utilize algorithms such as Decision Trees, Naïve Bayes, Random Forest, and SVM to classify diseases based on patient symptoms. These models are widely applied due to their interpretability and efficiency in structured data analysis. Table 1 summarizes studies that utilized predictive models for disease prediction.

Table 1. Predictive Model Comparison

Author	Technique	Dataset	Value of Measurement	Findings
A. Kumar Pandey et al. (2013)	Decision Tree	UCI Machine Learning Repository	Accuracy	J48 Pruned Decision Tree achieved 75.73% accuracy.
Tetiana Dudkina et al. (2021)	Decision Tree	Open database of 768 diabetes patients	Accuracy	Achieved 71% accuracy with a 50:50 data split for training/testing.
Chandrasekhar Rao Jetti et al.	Naïve Bayes	User-entered symptoms	Accuracy	Predicted disease and recommended doctors based on symptoms.
Mohammed Khalilia et al. (2011)	Random Forest	Nationwide Inpatient Sample (NIS)	Area under the ROC curve (AUC)	Random Forest achieved an average AUC of 88.79%.
Madhumita Pal and Smita Parija	Random Forest	Dataset from Kaggle	Accuracy, Sensitivity, Specificity	Achieved 86.9% accuracy and high sensitivity (90.6%) for heart disease.

2.2 Adaptive Models

Adaptive models like LSTM and RNN are highly effective for sequential and time-series data. These models are particularly useful for analyzing data that involves temporal dependencies. Table 2 provides an overview of studies using adaptive models.

Table 2. Adaptive Model Comparison

Author	Technique	Dataset	Value of Measurement	Findings
Xin Hong et al. (2019)	LSTM	MRI data from ADNI	AUC/mAUC	Achieved an average AUC of 89.05%.
Kuang Junwei et al. (2019)	LSTM	Time-series data	AUC/mAUC	Best performance with AUC/mAUC of 93.5%/77.7%.
Hadeel Ahmed Abd El Aal et al. (2021)	RNN-LSTM	ADNI datasets	Accuracy, Precision, F-Score, MCC	Achieved 95.8% accuracy and MCC=92.04% on DS1 dataset.
B. Sankara Babu et al. (2018)	GFMMNN + GWO + RNN	UCI datasets	Accuracy	Achieved 98.23% accuracy on Cleveland dataset.

2.3 Hybrid Predictive Models

Hybrid predictive models combine multiple algorithms to leverage their strengths and minimize weaknesses. Ensemble methods and model stacking are common techniques. Table 3 highlights studies employing hybrid models.

Table 1. Hybrid Predictive Model Comparison

Author	Technique	Dataset	Value of Measurement	Findings
Zhenya and Zhang (2021)	Ensemble Method	Mixed datasets	Accuracy	Ensemble of five classifiers significantly improved accuracy.
Das et al. (2009)	Stacked Ensemble	Heart disease dataset	Weighted-average voting	Combined Random Forest, SVM, and KNN predictions for better classification.

2.4 Related Work

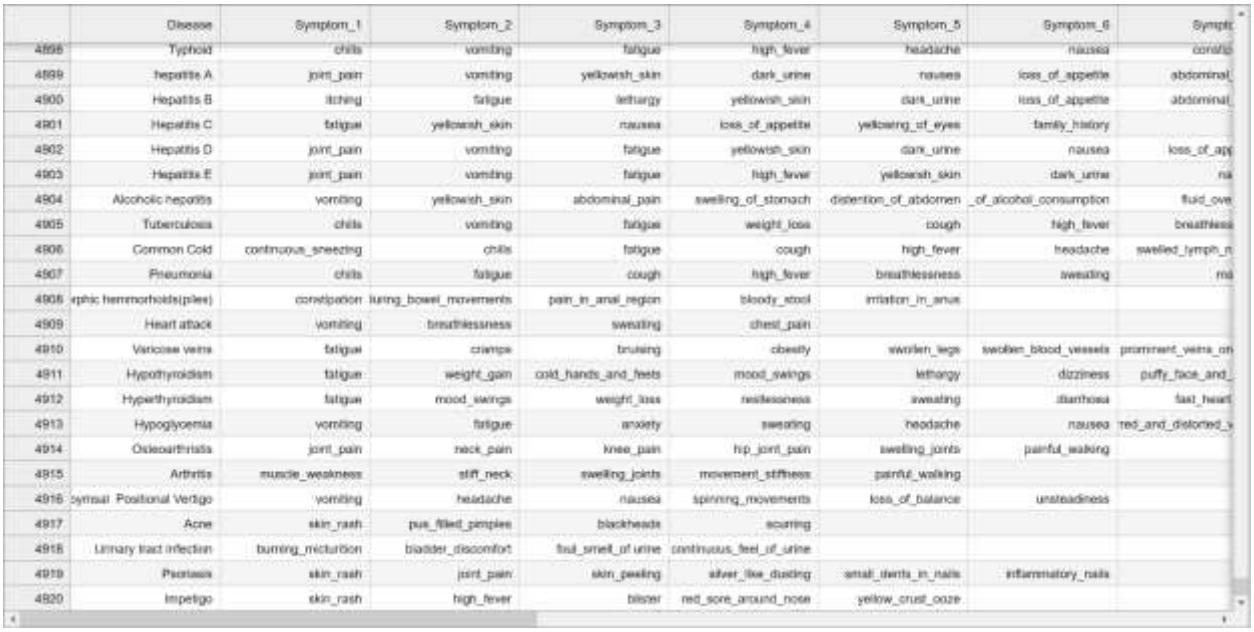
The evolution of disease prediction systems has seen significant advancements with the introduction of machine learning techniques. However, earlier studies often relied on standalone models, which, despite their simplicity and interpretability, faced limitations in terms of accuracy, robustness, and adaptability to diverse datasets. These limitations underscored the need for more sophisticated methods, such as hybrid models, which combine multiple algorithms to leverage their complementary strengths. Hybrid approaches not only address the constraints of standalone models but also introduce techniques like ensemble learning and model stacking to improve prediction performance. Table 4 provides a comparative overview of notable studies in the field, highlighting their methodologies, strengths, and areas for improvement.

Table 2. Related Work Comparison

Author	Methodology	Strengths	Limitations
Rahul Patil et al. (2016)	Hybrid model combining k-means clustering and Naïve Bayes classification	Focused on early diagnosis of diseases	Did not consider the severity of symptoms or integration with real-time patient data.
Archana L. Rane (2019)	Clinical decision support system using Decision Trees, Naïve Bayes, and SVM	Supported diagnostic decision-making	Lacked the implementation of ensemble methods to enhance accuracy.
Zhenya & Zhang (2021)	Cost-sensitive ensemble combining Random Forest, SVM, and K-Nearest Neighbor	Improved prediction accuracy through stacking	Limited to heart disease prediction; generalizability to other diseases was not explored in the study.

3.0 Research Methods

The dataset combines structured information from Kaggle, including patient demographics, symptoms, and diagnosed diseases. A supplementary dataset maps symptoms to severity scores, enhancing prediction accuracy, as shown in Figure 1.

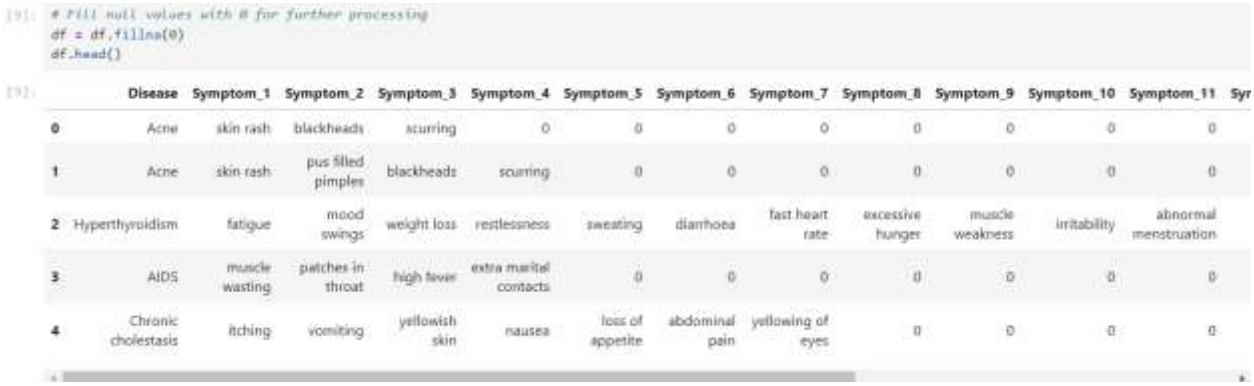


	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Sympt
4898	Typhoid	chills	vomiting	fatigue	high_fever	headache	nausea	constip
4899	Hepatitis A	joint_pain	vomiting	yellowish_skin	dark_urine	nausea	loss_of_appetite	abdominal_
4900	Hepatitis B	itching	fatigue	lethargy	yellowish_skin	dark_urine	loss_of_appetite	abdominal_
4901	Hepatitis C	fatigue	yellowish_skin	nausea	loss_of_appetite	yellowing_of_eyes	family_history	
4902	Hepatitis D	joint_pain	vomiting	fatigue	yellowish_skin	dark_urine	nausea	loss_of_ap
4903	Hepatitis E	joint_pain	vomiting	fatigue	high_fever	yellowish_skin	dark_urine	na
4904	Alcoholic hepatitis	vomiting	yellowish_skin	abdominal_pain	swelling_of_stomach	distention_of_abdomen	_of_alcohol_consumption	fluid_ove
4905	Tuberculosis	chills	vomiting	fatigue	weight_loss	cough	high_fever	breathless
4906	Common Cold	continuous_sneezing	chills	fatigue	cough	high_fever	headache	swelled_lymph_n
4907	Pneumonia	chills	fatigue	cough	high_fever	breathlessness	sweating	ma
4908	xphic hemmorhoids(piles)	constipation	luring_bowel_movements	pain_in_anal_region	bloody_stool	irritation_in_anus		
4909	Heart attack	vomiting	breathlessness	sweating	chest_pain			
4910	Varicose veins	fatigue	cramps	bruising	obesity	swollen_legs	swollen_blood_vessels	prominent_veins_on
4911	Hypothyroidism	fatigue	weight_gain	cold_hands_and_feets	mood_swings	lethargy	dizziness	puffy_face_and_
4912	Hyperthyroidism	fatigue	mood_swings	weight_loss	restlessness	sweating	diarrhoea	fast_heart
4913	Hypoglycemia	vomiting	fatigue	anxiety	sweating	headache	nausea	red_and_distorted_v
4914	Osteoarthritis	joint_pain	neck_pain	knee_pain	hip_joint_pain	swelling_joints	painful_walking	
4915	Arthritis	muscle_weakness	stiff_neck	swelling_joints	movement_stiffness	painful_walking		
4916	symoal Positional Vertigo	vomiting	headache	nausea	spinning_movements	loss_of_balance	unsteadiness	
4917	Acne	skin_rash	pus_filled_pimples	blackheads	scurring			
4918	Urinary tract infection	burning_micturition	bladder_discomfort	foul_smell_of urine	continuous_feel_of_urine			
4919	Psoriasis	skin_rash	joint_pain	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	
4920	Impetigo	skin_rash	high_fever	blister	red_sore_around_nose	yellow_crust_ooze		

Fig. 1 Dataset

3.1 Data Preprocessing

To ensure the dataset's quality and avoid potential issues during analysis, any null values in the data were replaced with 0. This approach helps maintain consistency across the dataset, allowing the machine learning algorithms to process the information effectively without encountering errors due to missing values. In addition, ambiguous entries—such as incomplete or contradictory data points—were carefully reviewed and removed to prevent confusion and ensure the accuracy of the predictions. This step is crucial in refining the dataset for optimal model performance. The data cleaning process is visually represented in Figure 2, which illustrates the steps taken to handle null and ambiguous values.



```
[9]: # Fill null values with 0 for further processing
     df = df.fillna(0)
     df.head()
```

[9]:

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7	Symptom_8	Symptom_9	Symptom_10	Symptom_11	Sy
0	Acne	skin rash	blackheads	scurring	0	0	0	0	0	0	0	0	
1	Acne	skin rash	pus filled pimples	blackheads	scurring	0	0	0	0	0	0	0	
2	Hyperthyroidism	fatigue	mood swings	weight loss	restlessness	sweating	diarrhoea	fast heart rate	excessive hunger	muscle weakness	irritability	abnormal menstruation	
3	AIDS	muscle wasting	patches in throat	high fever	extra marital contacts	0	0	0	0	0	0	0	
4	Chronic cholestasis	itching	vomiting	yellowish skin	nausea	loss of appetite	abdominal pain	yellowing of eyes	0	0	0	0	

Fig. 2 Remove Null values

To enhance the accuracy of the disease prediction model, symptoms were replaced with severity weights derived from a dedicated severity dataset. This dataset assigns a weight to each symptom based on its

perceived severity, which helps create a more personalized and reliable prediction model. Instead of simply using raw symptom data, the severity weights enable the model to prioritize more critical symptoms, reflecting their higher influence on the disease outcome. This process of assigning severity weights is visually depicted in Figure 3, which shows how symptom data is transformed into weighted values for use in the model.

```
# Replacing symptoms in 'df' with weights based on severity data in 'df1'
vals = df.values
symptoms = df1['Symptom'].unique()

for i in range(len(symptoms)):
    vals[vals == symptoms[i]] = df1[df1['Symptom'] == symptoms[i]]['weight'].values[0]

d = pd.DataFrame(vals, columns=cols)
d.head()
```

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7	Symptom_8	Symptom_9	Symptom_10	Symptom_11	Sy
0	Acne	3	2	2	0	0	0	0	0	0	0	0	
1	Acne	3	2	2	2	0	0	0	0	0	0	0	
2	Hyperthyroidism	4	3	3	5	3	6	5	4	2	2	6	
3	AIDS	3	6	7	5	0	0	0	0	0	0	0	
4	Chronic cholestasis	1	5	3	5	4	4	4	0	0	0	0	

Fig. 3 Using Severity Weight

3.2 Model Development

Selecting the right machine learning algorithms is crucial for building a robust and reliable disease prediction model. In this research, three algorithms were chosen for their unique strengths and complementary features, which make them well-suited for symptom-based disease prediction in healthcare:

Gradient Boosting Classifier (GBC):
Gradient Boosting is an ensemble method that builds models sequentially, with each iteration correcting the errors of the previous one. Its ability to minimize overfitting and handle complex decision boundaries makes it particularly effective for structured healthcare data. GBC is highly efficient for tasks requiring fine-grained decision-making, which is essential in accurately predicting diseases based on symptoms.

Decision Tree Classifier (DT):
Decision Trees are intuitive and interpretable models that classify data using hierarchical rules. Their simplicity and ease of use make them ideal for generating initial predictions and understanding key contributing features. As a foundational algorithm, Decision Trees are computationally efficient and integrate seamlessly into ensemble learning frameworks.

Random Forest Classifier (RFC):
Random Forest, a bagging method, constructs multiple Decision Trees and aggregates their predictions for improved accuracy. Known for its robustness against overfitting and ability to handle high-dimensional datasets, RFC is particularly effective in identifying feature importance, helping to uncover correlations between symptoms and diseases.

Justification for Algorithm Selection:
These algorithms were selected for their complementary strengths. GBC delivers high accuracy for structured data, RFC provides stability and robustness, and DT offers simplicity and interpretability. Together, these models form a hybrid approach that addresses the limitations of standalone algorithms, ensuring a reliable and efficient disease prediction system.

3.3 Evaluation Metrics

Models were assessed using accuracy, F1-score, and confusion matrices to evaluate their performance. Accuracy provides a straightforward measure of the model's overall correctness, while the F1-score balances precision and recall, making it particularly useful in healthcare settings where false positives and false negatives can have significant consequences. The confusion matrix offers a detailed breakdown of the model's classification performance, showing the true positives, true negatives, false positives, and false negatives. To ensure the reliability of the evaluation, cross-validation was performed, which involves splitting the data into multiple subsets and testing the model on different folds to assess its generalization ability. For a deeper understanding of the evaluation metrics and how they were calculated, refer to the evaluation metrics **formulas** presented in Figure 4.

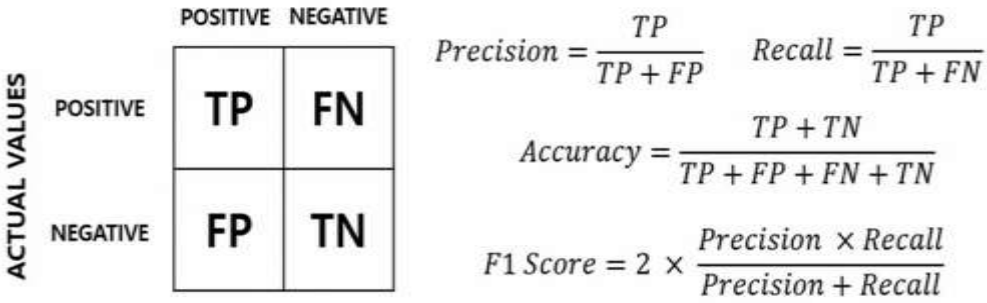


Fig. 4 Evaluation Metrics Formula

3.4 Hybrid Method (Voting Ensemble Method)

The Voting Ensemble method integrates the strengths of multiple algorithms to improve overall accuracy and reduce variance. By employing hard voting, this study combines predictions from GBC, DT, and RFC. The Voting Classifier aggregates the majority votes, ensuring that the final prediction benefits from the unique strengths of each model. This approach is particularly effective in healthcare, where minimizing prediction errors is critical. Studies show that ensemble methods outperform individual algorithms by leveraging diverse perspectives and reducing overfitting, making the Voting Ensemble an ideal choice for disease prediction. The concept of the Voting Ensemble method is illustrated in **Figure 5**, which visually represents how the predictions from multiple models are combined to form the final result.

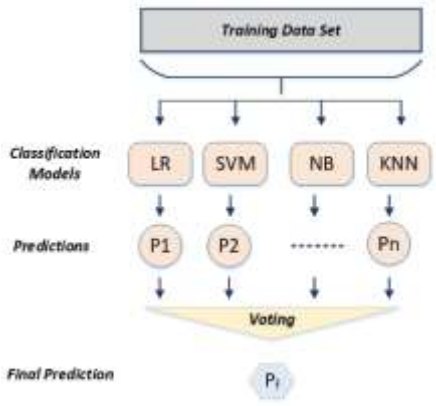


Fig. 5 Voting Ensemble

3.5 System Design

The system design focuses on integrating a hybrid machine-learning model with a user-friendly interface to predict diseases based on patient symptoms. The user interface is designed for simplicity, allowing users to easily input symptoms and receive disease predictions along with descriptions and precautions. By combining the power of machine learning with an accessible UI, the system ensures an intuitive experience while delivering accurate, actionable health insights, as shown in Figure 6.

Fig. 6 User Interfaces Design

4.0 Result and Discussion

In terms of descriptive statistics, the dataset consists of 133 unique symptoms mapped to multiple diseases, along with corresponding precautions. The severity of each symptom was accounted for by assigning severity weights, ensuring that critical symptoms were given higher importance in the prediction model. This approach allowed for more accurate disease predictions, as the model could focus on the most impactful symptoms, as shown in Figure 7, which illustrates the descriptive statistics of the dataset.

```
# Displaying count of unique symptoms and diseases
print("Symptoms used to identify the disease ",len(df1['Symptom'].unique()))
print("Diseases that can be identified ",len(df['Disease'].unique()))
```

```
Symptoms used to identify the disease  132
Diseases that can be identified  41
```

```
df['Disease'].unique()
```

```
array(['Acne', 'Hyperthyroidism', 'AIDS', 'Chronic cholestasis',
       'Hypertension', 'Hypoglycemia', 'Arthritis', 'Hepatitis B',
       'Migraine', 'Urinary tract infection', 'Diabetes', 'Hepatitis D',
       'Psoriasis', 'Alcoholic hepatitis', 'Dimorphic hemmorhoids(piles)',
       'Hepatitis E', 'Cervical spondylosis', 'Bronchial Asthma',
       'hepatitis A', 'Allergy', 'Hepatitis C', 'Pneumonia',
       'Hypothyroidism', 'Gastroenteritis', 'Varicose veins', 'Jaundice',
       'Drug Reaction', '(vertigo) Paroymsal  Positional Vertigo',
       'Heart attack', 'Tuberculosis', 'Typhoid', 'Common Cold',
       'Peptic ulcer disease', 'Paralysis (brain hemorrhage)',
       'Fungal infection', 'Impetigo', 'GERD', 'Dengue', 'Malaria',
       'Chicken pox', 'Osteoarthristis'], dtype=object)
```

Fig. 7 Descriptive Statistic

For model evaluation, the Gradient Boosting, Random Forest, and Voting Ensemble methods achieved an impressive accuracy of approximately 99.59%. The Voting Ensemble proved effective by combining the strengths of multiple algorithms. Although the Decision Tree model had a lower accuracy (~87.09%), its inclusion in the ensemble added diversity, contributing to the robustness of the final prediction. This demonstrates the power of ensemble methods in improving prediction performance by reducing variance and leveraging complementary strengths, as shown in Figure 8, which presents the accuracy results of all models.

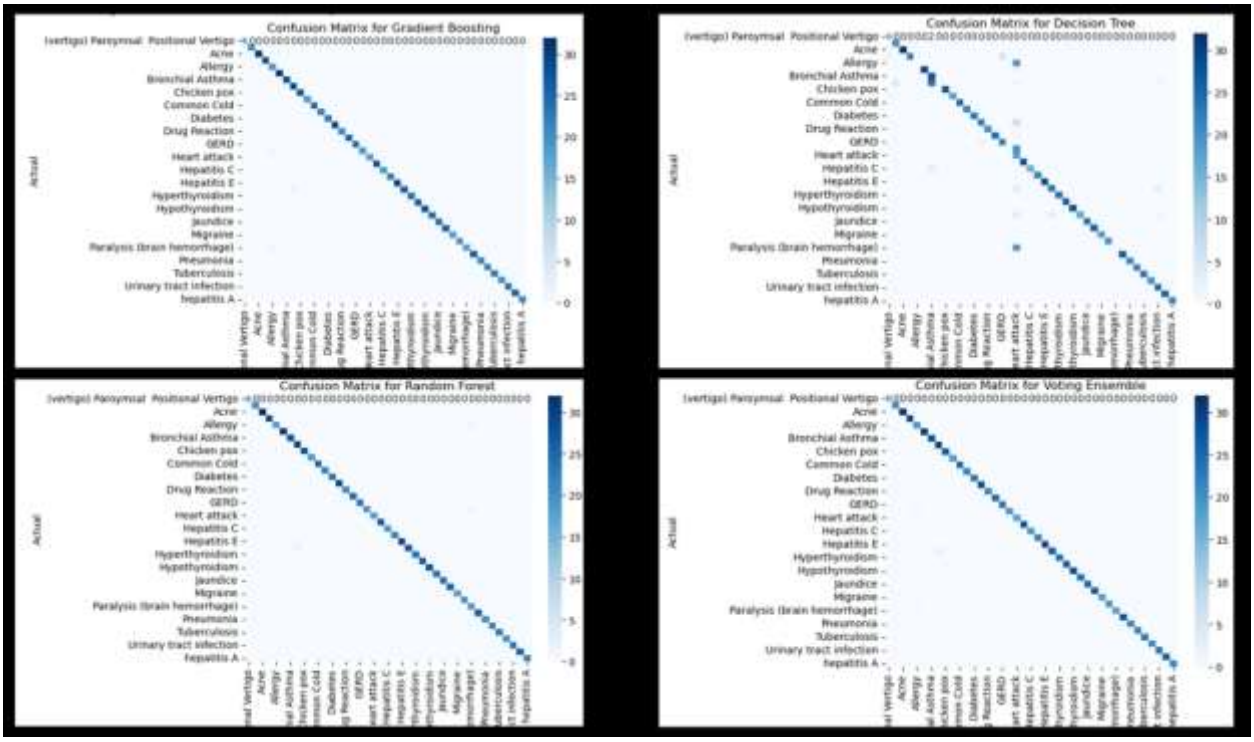


Fig. 8 Accuracy Results

5.0 Conclusions

The SHIELD system demonstrates the potential of hybrid machine learning models in advancing predictive healthcare. It achieves high accuracy and practical usability by leveraging ensemble techniques and incorporating symptom severity. Future research could focus on expanding the dataset and integrating NLP for unstructured symptom inputs, further enhancing system capabilities.

5. References

Gomathy, C. K., & Naidu, M. A. R. (2021). The prediction of disease using machine learning. International Journal of Scientific Research in Engineering and Management (IJSREM), 5(10), 1-7.

Kumar, P., & Shah, S. K. (2021). Disease Prediction Using Machine Learning Algorithms.

Sharma, V., Yadav, S., & Gupta, M. (2020, December). Heart disease prediction using machine learning techniques. In 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN) (pp. 177-181). IEEE.

Nayak, S. K., Garanayak, M., & Swain, S. K. (2024). A predictive prototype for the identification of diseases relied on the symptoms described by patients. EAI Endorsed Transactions on Pervasive Health and Technology, 10.

Takke, K., Bhaijee, R., Singh, A., & Patil, M. A. (2022). Medical disease prediction using machine learning algorithms. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10(V).

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. Ieee Access, 5, 20590-20616.

Joshi, K., Gupta, V. K., Jain, P., Shukla, A., Bharti, M., & Jindal, H. (2024, January). Disease Prediction System in Human Beings using Machine Learning Approaches. In 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-5). IEEE.

Subbalakshmi, G., Ramesh, K., & Rao, M. C. (2011). Decision support in heart disease prediction system using naive bayes. *Indian Journal of Computer Science and Engineering (IJCSE)*, *2*(2), 170-176.

Jetti, C. R., Shaik, R., Shaik, S., & Sanagapalli, S. (2021). Disease prediction using Naïve Bayes—machine learning algorithm. *International Journal of Science & Healthcare Research*.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.

Pal, M., & Parija, S. (2021, March). Prediction of heart diseases using random forest. In *Journal of Physics: Conference Series* (Vol. 1817, No. 1, p. 012009). IOP Publishing.

Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC medical informatics and decision making*, *11*, 1-13.

Pillai, N. S. R., Bee, K. K., & Kiruthika, J. (2019). Prediction of heart disease using rnn algorithm. *International Research Journal of Engineering and Technology*, *5*.

Abd El Aal, H. A., Taie, S. A., & El-Bendary, N. (2021). An optimized RNN-LSTM approach for parkinson's disease early detection using speech features. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2503-2512.

Che, C., Xiao, C., Liang, J., Jin, B., Zho, J., & Wang, F. (2017, June). An rnn architecture with dynamic temporal matching for personalized predictions of parkinson's disease. In *Proceedings of the 2017 SIAM international conference on data mining* (pp. 198-206). Society for Industrial and Applied Mathematics.

Vijayarani, S., Dhayanand, S., & Phil, M. (2015). Kidney disease prediction using SVM and ANN algorithms. *International Journal of Computing and Business Research (IJCBR)*, *6*(2), 1-12.

Sinha, P., & Sinha, P. (2015). Comparative study of chronic kidney disease prediction using KNN and SVM. *International Journal of Engineering Research and Technology*, *4*(12), 608-12.

Patil, R., Chopade, P., Mishra, A., Sane, B., & Sargar, Y. (2016). Disease prediction system using data mining hybrid approach. *Commun. Appl. Electron*, *4*.

Tarawneh, M., & Embarak, O. (2019). Hybrid approach for heart disease prediction using data mining techniques. In *advances in internet, data and web technologies: the 7th international conference on emerging internet, data and web technologies (EIDWT-2019)* (pp. 447-454). Springer International Publishing.

Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.

Junwei, K., Yang, H., Junjiang, L., & Zhijun, Y. (2019). Dynamic prediction of cardiovascular disease using improved LSTM. *International Journal of Crowd Science*, *3*(1), 14-25.

Hong, X., Lin, R., Yang, C., Zeng, N., Cai, C., Gou, J., & Yang, J. (2019). Predicting Alzheimer's disease using LSTM. *Ieee Access*, *7*, 80893-80901.

Parvez, S., Zubair, S., & Khan, A. (2023). A Hybrid Approach for Weak Learners Utilizing Ensemble Technique for Alzheimer's Disease Prognosis. *Indian Journal of Science and Technology*, *16*(32), 2518-2533.

Zhenya, Q., & Zhang, Z. (2021). A hybrid cost-sensitive ensemble for heart disease prediction. *BMC medical informatics and decision making*, *21*, 1-18.

Pandey, A. K., Pandey, P., Jaiswal, K. L., & Sen, A. K. (2013). A heart disease prediction model using decision tree. *IOSR Journal of Computer Engineering (IOSR-JCE)*, *12*(6), 83-86.

Dudkina, T., Meniailov, I., Bazilevych, K., Krivtsov, S., & Tkachenko, A. (2021). Classification and Prediction of Diabetes Disease using Decision Tree Method. In *IT&AS* (pp. 163-172).

Finlayson, T. L., Moyer, C. A., & Sonnad, S. S. (2004). Assessing symptoms, disease severity, and quality of life in the clinical context: a theoretical framework. *Am J Manag Care*, *10*(5), 336-344.