Predictive Modelling of Allergic Reactions Based on Lifestyle and External Triggers

Abhisha D Hedge
Department of CSE
NMAM Institute of Technology
Nitte (Deemed to be University)
Nitte - 574110, Karnataka, India
abhishahegde1@gmail.com

Diya R Kamath
Department of CSE
NMAM Institute of Technology
Nitte (Deemed to be University)
Nitte - 574110, Karnataka, India
diyakamath16@gmail.com

Harshith Durgadas Shettigar
Department of CSE
NMAM Institute of Technology
Nitte (Deemed to be University)
Nitte - 574110, Karnataka, India
hdharshith24@gmail.com

P Shivaram Aithal
Department of CSE
NMAM Institute of Technology
Nitte (Deemed to be University)
Nitte - 574110, Karnataka, India
shivaramaithal714@gmail.com

Shabari Shedthi B
Department of CSE
NMAM Institute of Technology
Nitte (Deemed to be University)
Nitte - 574110, Karnataka, India
shabarishetty@nitte.edu.in

***Abstract*—The reactions are quite variable among different individuals, depending on factors of lifestyle, exposure, and phys- iological conditions. Prediction of the level of severity of allergic reactions can greatly impact preventive and medical care against emergencies. The present project proposes a machine learning based system for the prediction of the severity of allergies from Dust, Pets, and Pollen based on patient lifestyle and symptoms.Various ML models were tested. The best accuracy was achieved by Stacking Ensemble along with Hyperparameter Tuning using RandomizedSearchCV. The maximum accuracy attained was up to 98.48.**

***Index Terms*—Allergy Severity Classification, Supervised Machine Learning, Stacking Ensemble Architecture, Feature Scaling and Encoding, Lifestyle–Symptom Feature Engineering, Hyperparameter Optimization (RandomizedSearchCV), Clinical Decision Support Systems.**

I. Introduction

Allergic reactions are among the most prevalent chronic ailments that affect millions of people all over the world. The effect of allergic reactions differs significantly from one person to another. The major factors that cause these reactions include the existence of dust, pollen, or fur of animals, along with personal attributes such as the quality of food, sleeping habits, exercise habits, smoking habits, and Body Mass Index.

Traditionally, the severity level of allergies is usually determined by clinical practice and the report provided by the affected individual. While these techniques are common, they are usually tedious, leading to inconsistencies in determining the level of severity. Moreover, the lack of predictive algorithms affects early detection and prevention, consequently rising in cases where severe allergic reactions are left unchecked due to lack of control. With the rising variables, manual interpretation becomes problematic.

But recent developments in the area of machine learning have made reliable and scalable approaches towards healthcare analytics possible. Machine learning models would be very useful in dealing with multi-dimensional data and deriving non-linear mappings from lifestyle variables to allergic reactions. Machine learning models use historical data to provide consistent and data-driven outputs that have little scope for human error in severity rating.

To address those challenges, the proposed solution seeks to develop and design a predictive system based on machine learning to predict the severity of allergies based on lifestyle, the nature of the allergies, and the allergy symptoms shown. This predictive system targets three forms of allergies, namely dust allergies, pollen allergies, and pet allergies, and gives the levels of allergy severity based on a predetermined scale of mild to very severe.

The main contributions of the proposed system are concluded in the following:

- **Lifestyle-Based Prediction:** Application of lifestyle features like diet score, sleeping patterns, physical activity levels, smoking status, and BMI for better prediction of severity.
- **Multi-Model Evaluation:** The process of evaluating the performance of different machine learning models to select the suitable models for predictions. It helps in determining.
- **Ensemble Learning:** Use of hyperparameter optimization methodologies for improving generalization capabilities and avoiding overfitting.
- **Model Optimization:** Use of hyperparameter optimization methodologies for improving generalization capabilities and avoiding overfitting.

The system proposed here attempts to create a reliable, automated, and scalable platform for predicting the severity of allergies. The system, which reduces subjectivity and allows early risk identification, can potentially contribute to preventive healthcare and help medical professionals with informed decision-making.

II. Literature Survey

This section discusses the current research work available in relation to predicting allergy risk and severity, machine learning in healthcare, lifestyle health modeling, and ensemble learning methods. The aim is to examine the current methods used and determine why the proposed system is required.

A. Allergy Risk and Severity Prediction Models

here have been a few studies that have investigated the application of predictive modeling for evaluating the risk and severity of allergies. The initial studies were based on statistical and mathematical models, using clinical parameters and patient history. Pattarakiatjaroen et al. have suggested a clinical prediction model for evaluating the risk of allergy using statistical modeling, emphasizing the difficulties involved in dealing with variability among patient populations [1]. Similarly, Santurino et al. have described mathematical models for predicting the risk of drug allergy, but emphasized the difficulties involved in scalability and adaptability [3].

More recent research has been directed at structured risk assessment models, especially in pediatric and chronic allergy patients. Namazova-Baranova et al. highlighted the value of early risk stratification in pediatric allergic diseases, although they also noted the subjective nature of risk stratification in conventional methods [8]. Analysis of computational models for predicting allergies also suggests that most current models are not generally applicable to different levels of allergy severity [15].

B. Machine Learning in Healthcare and Allergy Diagnosis

The integration of machine learning algorithms in the healthcare sector has improved the diagnosis and risk of diseases. Yang et al. proved the application of machine learning in large-scale allergy surveillance by analyzing clinical texts, which showed better detection abilities than traditional methods [4]. Likewise, Park et al. used machine learning algorithms to predict the development of allergic diseases, which showed better predictive abilities than traditional methods [7].

The use of AI-assisted clinical decision support systems has also been investigated for the purpose of assisting with allergy diagnosis and risk evaluation. The application of machine learning-based decision support systems in reducing variability in diagnosis was emphasized by Boyd et al. in their work, which showed the potential of machine learning models in dealing with complex healthcare data, while also indicating the issues associated with robustness and interpretability.

C. Lifestyle-Based Health and Allergy Prediction

Lifestyle and environmental factors are important in the development and progression of allergic diseases. Many studies have included lifestyle variables such as exercise, sleep quality, and environmental exposure in predictive models. Yenra showed that the inclusion of environmental and behavioral variables in AI models increases the accuracy of predictions for allergies [5]. Similarly, large studies on asthma and the prediction of risk of allergic diseases have highlighted the importance of prenatal, postnatal, and environmental factors in the progression of the disease [21], [22].

However, despite these improvements, many lifestyle-based models are centered on the onset of disease rather than the classification of severity. Furthermore, the analysis of lifestyle variables is often conducted individually or with minimal clinical information, making them less effective in a full allergy severity prediction system.

D. Ensemble Learning and Advanced Predictive Techniques

Ensemble learning algorithms have also been explored for their potential in enhancing the accuracy of predictions and the stability of models by using a combination of classifiers. The efficacy of ensemble learning in predicting the results of oral food challenges was demonstrated by Zhang et al., who found it to be more accurate than individual models [23]. Ensemble learning algorithms are effective in reducing bias and variance, which makes them suitable for handling complex medical data.

Deep learning techniques have also been used for prediction tasks related to allergies. Liu et al. introduced deep learning models for the prediction of allergenicity based on biological sequence information, and Yu et al. investigated multimodal deep learning models for the prediction of allergenic proteins [18], [19]. Although these models are highly accurate, they tend to be computationally expensive and require a substantial amount of data, making them less suitable for clinical prediction systems based on lifestyle.

E. Analysis of Research Gaps

Based on the literature reviewed, some gaps in the existing literature can be identified. Firstly, most of the existing literature is concerned with the risk of allergy or the onset of the disease, but not multi-level predictions of the severity of the disease. Secondly, lifestyle variables are not fully explored or incorporated with symptom data. Thirdly, although ensemble learning has been promising, its use in predicting the severity of an allergy is still limited. Lastly, most of the literature does not stress the importance of hyperparameter optimization.

The proposed project fills the gaps by designing a lifestyle-aware system for predicting the severity of allergies using a stacking ensemble learning approach. The system will be able to provide accurate results by combining the predictions of various base learners with a meta-learner.

III. Methodology

This section outlines the general methodology used to create the allergy severity prediction system. The proposed method uses a structured machine learning process that includes data collection, processing, model building, hyperparameter tuning, and model evaluation. The emphasis is on creating a reliable and accurate predictive model using ensemble learning methods.

A. Data Collection and Feature Selection

The dataset for this analysis contains patient attributes that include demographic information, lifestyle, allergy type, and symptoms. The features selected for this analysis are age, body mass index (BMI), sleep quality, diet quality, physical activity level, smoking status, allergy type, and symptoms. The features were selected based on their known effects on the immune system and allergy severity. The target variable is the level of allergy severity, which is measured on a five-point scale from mild to very severe.

B. Data Preprocessing

Data preprocessing was carried out to make the data consistent and amenable to machine learning algorithms. Missing values were treated using relevant statistical imputation methods. Categorical variables like allergy type and symptoms were converted to numerical variables using label encoding. Numerical variables were scaled using StandardScaler to make all variables uniformly scaled. The preprocessed data was then ready for model training and evaluation.

C. Stacking Ensemble Model Architecture

The proposed system uses a stacking ensemble architecture to enhance the predictive capability and robustness of the system. The stacking ensemble technique is a method that combines the predictions of different base learners and uses a meta-learner to learn how to combine the predictions.

The ensemble architecture is made up of the following components:

1) **Base Learners:**
   - Random Forest (tuned)
   - Gradient Boosting (tuned)
   - Logistic Regression (tuned)
2) **Meta Learner:**
   - Logistic Regression

Each base learner makes intermediate predictions, which are used as input features for the meta-learner. The meta-learner makes the final prediction of allergy severity.

D. Hyperparameter Optimization

Hyperparameter optimization was carried out to improve the performance of the models and prevent overfitting. The use of RandomizedSearchCV was implemented to effectively search the hyperparameter space of the base learners and the meta-learner. This method enables the random sampling of parameter settings, which strikes a balance between efficiency and performance improvement. Cross-validation was also implemented during the optimization process to accurately estimate the generalization performance of the models, including the minority class predictions.

E. Model Training and Validation

The preprocessed dataset was divided into training and testing sets. During the training process, each of the base learners was trained separately based on the training set. The predictions made by the base learners were combined to create meta-features, which were used as input for the meta-learner. The meta-learner was trained based on the meta-features to learn the best way of combining the predictions. Model validation was performed using cross-validation.

F. Performance Evaluation Metrics

The performance of the proposed model was evaluated using the following standard classification metrics: Accuracy, Precision, Recall, and F1-score. These metrics give a valid measurement of the model's predictions, particularly in handling imbalanced severity classes., emphasised on evaluating the model's ability to correctly identify higher severity levels, which are critical for applications like preventive healthcare applications.

G. System Workflow Overview

The entire workflow for the proposed system can therefore be described as follows:

- Input patient lifestyle variables, type of allergy, and symptoms.
- Preprocess the data and scale the features.
- Produce predictions using tuned base learners.
- Combine the outputs of multiple base learners using a meta-learner.
- Finalize the prediction for the allergy score.

In this way, this well-structured pipeline makes it possible to achieve accurate, automated, as well as scalable predictions of severity for allergies.

IV. Results and Discussion

This section highlights the experimental results acquired from the proposed system for allergy severity prediction and will analyze the results acquired from the stacking ensemble model. The results will be analyzed based on accuracy and effectiveness in classification.

A. Overall Model Performance

The finalized stacking ensemble model performed remarkably well on the predictive task for the tested dataset. The performance parameters of the developed system have been shown in Table I. The model exhibited an accuracy of 98.48%, thus signifying high correctness in predicting levels of allergy. Furthermore, precision, recall, and F1-score parameters were all in excess of 99%, thus signifying that the developed system predicts with reliability, taking into account both false positives and false negatives.

The high precision indicates that the model makes very few incorrect severity predictions, while the high recall shows its effectiveness in identifying true severity cases. The balanced F1 score further confirms the robustness of the model.

The performance metrics are presented in Table I.

TABLE I: Performance Metrics of the Proposed Model

Metric	Value (%)
Accuracy	98.48
Precision	99.33
Recall	99.00
F1-Score	99.10

B. Performance for Various Types of Allergies

This model has been tested individually for pet, pollen, and dust allergies to check its flexibility with different types of allergies. The accuracy of each type of allergy is shown in Table II. Out of the three types, the model gave the highest accuracy of 98.48% in dust allergies, followed by 96.55% in pollen allergies, and the last one was pet allergies with an accuracy of 96.25%.

The reasons for improved performance for dust allergy could be related to more stable symptom presentation patterns and stronger relationships between lifestyle variables and levels of severity. Notwithstanding differences in symptom presentation, it performed at a stable high level for all types of allergies, thus establishing its ability to generalize.

Table II presents the classification accuracy across different allergy types.

TABLE II: Final Accuracy Across Different Allergy Types

Allergy Type	Final Accuracy (%)
Pet Allergy	96.25
Pollen Allergy	96.55
Dust Allergy	98.48

C. Impact of Ensemble Learning and Hyperparameter Optimization

The stacking ensemble architecture proved to be one of the most important factors that increased the reliability of predictions. Tree models like Random Forest and Gradient Boosting were very effective at modeling non-linear relationships and interaction, and Logistic Regression also helped to ensure a strong boundary during decision-making. The Logistic Regression meta-learning model effectively learned to combine predictions from constituent models to produce better results.aly alerts generated by the system dashboard.

Hyperparameter tuning via RandomizedSearchCV added some model performance improvements by eliminating overfitting. The models were found to converge and predict better, especially when the severity level was higher, as it is a rare case in a medical dataset.

D. Discussion of Key Observations

The key observations made from the experimental evaluation are given below:

- This makes the stacking ensemble model very reliable and accurate in the modeled data consistently.
- LSR features contributed considerably toward effective severity discrimination.
- Hyperparameter tuning enhanced the stability and generalization of the model.
- The system performed well on different allergy categories without requiring any allergy-specific architecture changes.

These observations confirm that integrating ensemble learning with optimized model parameters provides a robust and scalable solution for allergy severity prediction.

E. Summary of Results

The proposed methodology and system have been successfully validated for its efficiency in solving the problem of determining the level of allergic severity using an ensemble method. The proposed system, which uses lifestyle factors and ensemble methods to predict hypersensitivity levels, has proven to be efficient in preventive healthcare.

V. Conclusion and Future Work

A. Conclusion

In this project, the system for predicting the intensity of allergic reactions based on lifestyle parameters, the type of allergy, and symptoms was developed using machine learning techniques. In the proposed system, various optimized machine learning models were combined into a stacking ensemble approach, which provided high accuracy and good generalization for various allergy types.

The combination of Random Forest, Gradient Boosting, and Logistic Regression methods in base learning, along with the meta-learning approach using Logistic Regression, helped in efficiently learning linear as well as non-linear patterns in the data. Hyperparameter tuning using RandomizedSearchCV increased the efficiency of the model by preventing it from being overfit. The results obtained were accurate to 98.48%, with complete precision, recall, and F1-score values, thus validating the efficiency of the proposed technique.

The results show that adding lifestyle attributes together with symptoms is a significant factor in improving predictions of allergy severity. Using this system, preventive health care, aid in decision-making tools for medical professionals, or an individualized risk of an allergy can be made scalable.

B. Future Work

Although this system shows good performance, there are some areas which need improvement in further work. Real-time information, such as pollen index, air quality, and weather, can be used to further improve the accuracy of this system. This system can also be enhanced to incorporate other types of allergies, such as food allergy or allergy to chemicals.

Future applications may consider taking advantage of deep learning methods, such as recurrent networks or those that employ an attention mechanism, in an effort to leverage time-related patterns of symptoms. Furthermore, a web or mobile application deployment would provide an opportunity to engage patients in real time, increasing availability. Moreover, clinician input might be considered in an attempt to adapt to evolving patterns in the data.

References

[1] S. Pattarakiatjaroen et al., "Clinical prediction of allergy risk using statistical modeling," 2025.

[2] A. DunnGalvin et al., "Predictive modeling approaches for food allergy diagnosis," 2025.

[3] I. Santurino et al., "Mathematical modeling for drug allergy risk prediction," 2016.

[4] Y. Yang et al., "Machine learning analysis of clinical text for allergy surveillance," *JAMA Network Open*, 2020.

[5] Yenra, "AI-based allergy prediction using genetics, family history, and environmental data," *Yenra AI Report*.

[6] Food Allergy Institute/TIP, "Integrated AI platform for early food allergy prediction," 2025.

[7] J. Park et al., "Machine learning-based prediction of allergic disease onset," 2024.

[8] L. Namazova-Baranova et al., "Risk assessment models for pediatric allergic diseases," 2024.

[9] R. Czolk et al., "Biomarker-driven prediction of allergic conditions," 2021.

[10] X. Wang et al., "Deep learning approaches for allergic disease classification," 2023.

[11] S. Boyd et al., "Clinical decision support systems for allergy risk assessment," 2024.

[12] A. Firza et al., "Machine learning applications in allergy diagnosis during pandemics," 2022.

[13] G. Russo et al., "AI-assisted screening for allergic disease risk," 2024.

[14] Asthma ML Study, "Machine learning-based prediction of asthma risk," 2025.

[15] M. van Breugel et al., "Review of computational models for allergy prediction," 2023.

[16] Kim et al., "Cutoff-based diagnostic modeling in allergic disease prediction," 2021.

[17] Zhao et al., "Data-driven approaches for allergy risk stratification," 2024.

[18] J. Liu, S. S. Negi, C. Y. et al., "AllergenAI: a deep learning model predicting allergenicity based on protein sequence," *BMC Bioinformatics*, vol. 26, art. no. 279, 2025.

[19] L. Yu, Y. Luo, S. Wu et al., "Multimodal deep learning for allergenic proteins prediction," *BMC Biology*, vol. 23, art. no. 232, 2025.

[20] B. Liu, Z. Yang, Q. Liu et al., "Computational prediction of allergenic proteins based on multi-feature fusion," *Frontiers in Genetics*, vol. 14, art. no. 1294159, Oct. 2023.

[21] H. Nagase et al., "Machine learning for prediction of asthma exacerbations among asthmatic patients: a systematic review and meta-analysis," *BMC Pulmonary Medicine*, 2023.

[22] Z. Jeddi et al., "Machine learning for predicting the risk for childhood asthma using prenatal, perinatal, postnatal and environmental factors," *Healthcare*, vol. 9, no. 11, 2021.

[23] X. Zhang, D. Lee, K. Jungles et al., "Prediction of oral food challenge outcomes via ensemble learning," *arXiv preprint*, 2022.

[24] Y. Li, Y. Xiong, W. Fan et al., "Sequential model for predicting patient adherence in subcutaneous immunotherapy for allergic rhinitis," *arXiv preprint*, 2024.

[25] L. Chen et al., "Temporal modelling of allergy progression using recurrent neural networks," *BMC Medical Informatics and Decision Making*, 2021.