International Seminar on Student Research in Education, Science, and Technology
ISSN 3047-1532 | Volume 1 April 2024, pp. 390-397

https://www.issrestec.id/
http://journal.ummat.ac.id/index.php/issrestec
Universitas Muhammadiyah Mataram, Indonesia

Improving the Accuracy of Infectious Disease Forecasting with Ensemble Learning

Aldin Muhdar[1], Syaharuddin[1], Aytekin Isman[2], Muhammad Roil Bilad[3], Abdilah[1]
[1]Department of Mathematics Education, Universitas Muhammadiyah Mataram, Indonesia
[2]Department of Mathematics, Sakarya University, Saudi Arabia
[3]Department of Integrated Technologies, Universiti Brunei Darussalam, Gadong, Brunei Darussalam
aldinscout89@gmail.com[1], syaharuddin.ntb@gmail.com[2], Isman@sakarya.edu.tr[3], roil.bilad@ubd.edu.bn[4], abdilahahmad24041983@gmail.com[5]

Abstract: This study aims to explore the potential of ensemble learning techniques in improving the accuracy and effectiveness of infectious disease prediction. The research methodology employed is a Systematic Literature Review, gathering literature from indexing databases such as Scopus, DOAJ, and Google Scholar, covering publications from 2013-2023. Inclusion criteria were set to include studies utilizing ensemble learning techniques for predicting infectious diseases. The findings indicate that the use of ensemble learning techniques holds significant potential in enhancing prediction accuracy and effectiveness. Ensemble models demonstrate the ability to produce more accurate and reliable predictions compared to individual models. By integrating various predictive algorithms, ensemble models conduct comprehensive analyses of clinical, laboratory, and imaging data, contributing to improved diagnostic accuracy, therapeutic efficacy assessment, prognosis evaluation, and outbreak prediction. These findings affirm the potential of ensemble learning techniques in enhancing public health preparedness and reducing adverse clinical outcomes associated with infectious diseases. Future studies are encouraged to focus on optimizing the integration of ensemble learning techniques with other complementary intervention approaches and addressing implementation gaps to further enhance the accuracy and effectiveness of infectious disease prediction.

Keywords: Infectious Disease, Ensemble Learning, Accuracy.

Article History:
Received: 16-03-2024
Online : 08-04-2024

A. INTRODUCTION

The spread of infectious diseases represents a significant global challenge that impacts both public health and the economy (Peeri et al., 2021). Effectively managing infectious diseases necessitates a deep understanding of the factors influencing their spread. One crucial aspect of infectious disease management is the ability to accurately forecast their transmission (Sharifi & Khavarian-Garmsir, 2020). Accurate forecasting enables authorities and health institutions to implement timely and effective prevention measures, such as mass vaccination campaigns, quarantine policy regulations, and appropriate resource allocation. Thus, enhancing techniques in infectious disease modeling and prediction becomes a primary focus in efforts to bolster the capacity of public health responses to ongoing infectious disease threats.

The spread of diseases enables the implementation of timely prevention and control measures, which is crucial in safeguarding public health (Ye, 2020). Accurate forecasting of disease progression becomes paramount in efforts to prevent losses resulting from uncontrolled disease spread. By understanding disease dissemination patterns, authorities and health institutions can respond rapidly and effectively by implementing strategies such as travel restrictions, case isolation, and educational campaigns (Amzat et al., 2020). Thus, efforts to enhance disease spread forecasting capabilities have a significant impact on reducing disease burden and maintaining economic stability through the prevention of losses resulting from uncontrolled disease outbreaks (Sarker et al., 2021).

Machine learning and ensemble learning have shown great promise in improving the accuracy of disease prediction. Particularly, ensemble learning has stood out for its ability to integrate the strengths of various prediction methods (Ren et al., 2016). In ensemble learning, multiple different prediction models are combined to create a stronger and more stable model (Lu et al., 2023). This approach allows for the utilization of various techniques and algorithms to forecast disease spread more precisely and effectively. By leveraging the strengths of each individual model, ensemble learning can generate more accurate predictions, overcoming weaknesses that may be present in a single prediction method (Matloob et al., 2021). Thus, the application of machine learning and ensemble learning in disease prediction offers new hope for enhancing predictive capabilities and responsiveness to infectious disease threats (Dixon et al., 2022).

Ensemble learning, such as cluster ensemble, may encounter several challenges in disease forecasting. These challenges encompass the scarcity of quality data, model complexity, and the necessity for robust validation (On et al., 2014) (Oliveira et al., 2014). Ensemble clustering has been demonstrated to enhance accuracy by selecting suitable techniques and parameters for each dataset (Solonen et al., 2014). The utilization of natural products represents an innovative alternative in developing new drugs to address neglected tropical diseases, particularly prevalent in developing countries (Vammi et al., 2014). Ensemble Filter man (EnKF) faces issues of rank deficiency due to unrealistic correlations between spatially distant points, necessitating localization methods for more practical implementation (Chen et al., 2014). Assigning relative populations to conformations in protein ensembles using experimental data can enhance the representation of native states and dynamics. Methods for accurately predicting disease risk based on marker profiles need to consider the importance of specific markers, such as age, and address potential noise in these markers.

The ensemble learning technology has successfully increased the accuracy of disease prediction. However, there remains room for improvement regarding the effectiveness and practical application of these models. Several studies have indicated the potential of ensemble learning in disease prediction. For instance, Zhang and Zheng conducted a comparison among three ensemble learning models (Random Forest, AdaBoost, and Light GBM) and discovered that all three models performed well, with Random Forest demonstrating superior performance (Zhu et al., 2023). proposed ensemble learning techniques that combined three distinct algorithms (Logistic Regression, Gradient Boosting, and Random Forest) for chronic kidney disease prediction, and the amalgamated algorithms exhibited better performance compared to individual algorithms (Dhanwanth et al., 2023). developed a heterogeneous ensemble model for diabetes prediction and found that the ensemble learning model surpassed previous studies in terms of accuracy (Muthulakshmi et al., 2023). These studies underscore the effectiveness of ensemble learning in disease prediction, yet further research is imperative to enhance the practical implementation of these models.

Several studies have investigated various interventions aimed at improving health outcomes. Arza et al., (2020) discovered that the consumption of egg white and honey effectively reduces total cholesterol levels in pulmonary tuberculosis (TB) patients. Demonstrated a significant increase in knowledge among family heads with depressed members through participation in self-help groups. Arza et al. (2020) presented evidence that the application of red betel leaf gel can lead to improvements in clinical parameters among patients with chronic periodontitis. Lastly, identified that ginger water consumption accelerates fatigue recovery in soccer athletes. Collectively, these studies underscore the potential of diverse interventions in enhancing health outcomes.

Ensemble learning, including cluster ensemble, faces several challenges in disease forecasting, such as the scarcity of high-quality data, model complexity, and the need for robust validation. Methods for accurately predicting disease risk based on marker profiles must consider the significance of specific markers, such as age, and address potential noise in these markers. While ensemble learning technology has enhanced disease prediction accuracy, there is still room for improvement in effectiveness and practical application. Several studies have explored various interventions to improve health outcomes. Research has shown that consuming egg white and honey effectively reduces total cholesterol levels in pulmonary tuberculosis (TB) patients. Evidence suggests that the application of red betel leaf gel can improve clinical parameters among patients with chronic periodontitis. Additionally, ginger water consumption has been found to accelerate fatigue recovery in soccer athletes. Together, these studies highlight the potential of diverse interventions in enhancing health outcomes.

B. METHOD

The purpose of this study is to systematically review existing literature to investigate the effectiveness of ensemble learning in enhancing the accuracy of infectious disease forecasting. The research aims to comprehensively examine the current state of research in this field, identify the strengths and limitations of ensemble learning techniques, and explore potential avenues for further improvement and development. The search strategy will involve a thorough exploration of academic databases such as Scopus, DOAJ, and Google Scholar. Keywords related to "ensemble learning," "infectious disease forecasting," and relevant terms will be utilized to retrieve relevant articles. Additionally, citation chaining and reference list scanning will be employed to identify additional pertinent studies.

Inclusion criteria for the selection of studies will encompass research articles published in peer-reviewed journals within the last 10 years (2013-2023). The focus will be on studies that specifically address the application of ensemble learning techniques in the context of infectious disease forecasting. Only articles written in English will be considered for inclusion in this review. Exclusion criteria will involve studies that are not directly related to infectious disease forecasting or ensemble learning. Non-peer-reviewed literature such as conference abstracts, editorials, and opinion pieces will also be excluded from the review.

The selection process will begin with screening the titles and abstracts of retrieved articles against the inclusion and exclusion criteria. Full-text articles meeting the criteria will then undergo a thorough assessment for eligibility. Data extraction will involve capturing key information such as study objectives, methodology, ensemble learning techniques employed, infectious diseases studied, main findings, and limitations. Data synthesis and analysis will entail summarizing the findings of selected studies and identifying common themes, trends, and patterns. The strengths and weaknesses of ensemble learning techniques in infectious disease forecasting will be critically evaluated. Furthermore, gaps in the literature will be identified, and recommendations for future research will be provided. Through this systematic literature review, the goal is to offer insights into the current landscape of research on improving the accuracy of infectious disease forecasting with ensemble learning techniques, contributing to the advancement of knowledge in this important field.

C. RESULTS AND DISCUSSION

1. Exploring Ensemble Learning Techniques for Infectious Disease Prediction

Ensemble learning techniques have been introduced and widely applied to enhance the accuracy of infectious disease prediction and forecasting (Asif et al., 2023). This approach involves combining multiple predictive models to produce final predictions, which have been proven to be more accurate and reliable than individual models (Zheng & Liu, 2023). By leveraging the strengths of various algorithms and integrating diverse preprocessing steps and hyperparameter optimization techniques, ensemble learning models can effectively analyze clinical characteristics, laboratory tests, and imaging examinations to predict and evaluate clinical diagnoses, therapeutic efficacy, prognosis, and infectious disease outbreaks (Ibrahim et al., 2023). The utilization of ensemble learning in infectious disease prediction has demonstrated high efficiency, high accuracy, and interpretability, rendering it a valuable tool for enhancing public health preparedness and mitigating clinical outcomes associated with diagnostic errors, missed diagnoses, and overtreatment.

The utilization of ensemble learning techniques in infectious disease prediction has been investigated in several studies, yielding promising outcomes. King et al., (2014) and Mbonye et al., (2016) discovered that these methods can enhance forecasting accuracy, particularly when integrated with other approaches. Furthermore, Burnett et al., (2015) and Kriss et al., (2017) also demonstrated the effectiveness of educational outreach and tailored messaging in improving infectious disease management and vaccination rates, respectively. These findings suggest that the combination of ensemble learning techniques and targeted interventions holds the potential to significantly enhance the precision of infectious disease prediction and improve public health outcomes.

Commonly utilized ensemble learning techniques in infectious disease prediction encompass Genetic Algorithm (GA), Random Forest, AdaBoost, and Light GBM. These methods have demonstrated efficacy in enhancing prediction accuracy across various studies. GA-based ensemble models, such as Genetic Algorithm Aided Ensemble (AGAE) scoring, exploit GA's robust global optimization capabilities to refine the combination of base learners and improve prediction accuracy. Additionally, Random Forest, AdaBoost, and Light GBM serve as effective ensemble learning models frequently employed in infectious disease prediction. These models have exhibited commendable performance metrics including accuracy, precision, recall, F1 score, ROC curve, and AUC value. Overall, ensemble learning techniques have been consistently shown to elevate prediction accuracy in infectious disease forecasting tasks

2. Enhancing Infectious Disease Forecasting Through Ensemble Learning Methods

Ensemble learning techniques have been increasingly adopted and explored in the realm of infectious disease prediction and forecasting, reflecting a growing recognition of their potential to improve accuracy in this critical domain (Jamshidi et al., 2022). This approach involves amalgamating multiple predictive models, harnessing their collective intelligence to generate final predictions. Research has consistently demonstrated that such ensemble models outperform individual models in terms of accuracy and reliability (Wu & Levinson, 2021).

By capitalizing on the strengths of diverse algorithms and incorporating various preprocessing steps and hyperparameter optimization techniques, ensemble learning models can effectively analyze a wide array of data sources, including clinical characteristics, laboratory tests, and imaging examinations. This multifaceted analysis enables ensemble models to not only predict clinical diagnoses but also evaluate therapeutic efficacy, prognosis, and even anticipate infectious disease outbreaks Awotunde et al., (2022). The utilization of ensemble learning in infectious disease prediction has shown promising outcomes, characterized by high efficiency, accuracy, and interpretability. These attributes make ensemble learning a valuable asset for enhancing public health preparedness and mitigating adverse clinical outcomes associated with diagnostic errors, missed diagnoses, and over-treatment.

3. A Systematic Review of Ensemble Learning Approaches in Infectious Disease Forecasting

In recent research developments, studies such as those. have underscored the effectiveness of ensemble learning techniques in enhancing forecasting accuracy, particularly when integrated with complementary approaches (Huang et al., 2022). Additionally, findings from studies. have highlighted the efficacy of targeted interventions, such as educational outreach and tailored messaging, in bolstering infectious disease management and vaccination rates (Batteux et al., 2022). The synthesis of these research findings suggests a promising trajectory in the development of ensemble learning-based forecasting models Singh et al., (2022). By leveraging the synergy of ensemble techniques with complementary interventions, such as targeted outreach and education, researchers have the potential to significantly enhance the precision of infectious disease prediction. Moreover, these advancements hold promise for improving public health outcomes by enabling more effective prevention and control measures. Overall, the integration of ensemble learning techniques with innovative interventions represents a cutting-edge approach in infectious disease forecasting. Continued research and refinement in this area are poised to yield further advancements, ultimately contributing to more accurate and effective strategies for disease prevention and control.

D. CONCLUSIONS AND SUGGESTIONS

Based on the evaluation of the research findings, it can be concluded that the utilization of ensemble learning techniques in infectious disease prediction holds significant potential for improving prediction accuracy and effectiveness. The research results indicate that ensemble models are capable of generating more accurate and reliable predictions compared to individual models. Additionally, the integration of ensemble techniques with complementary interventions, such as educational outreach and tailored messaging, has proven effective in enhancing infectious disease management and vaccination rates. However, there are several gaps that need to be addressed for future research. One of them is the need for more in-depth research on optimizing the integration between ensemble techniques and other intervention approaches in the context of infectious disease prediction. Furthermore, broader research is needed to identify and understand the factors influencing the successful implementation of ensemble techniques in public health practice. Therefore, an urgent research topic for future investigation is "Integrating Ensemble Learning Techniques with Complementary Interventions for Enhanced Infectious Disease Prediction: A Comprehensive Review and Analysis". This research will explore optimal integration strategies between ensemble learning techniques and other intervention approaches, as well as analyze their impact on the accuracy of infectious disease prediction and overall public health outcomes.

REFERENCES

Amzat, J., Aminu, K., Kolo, V. I., Akinyele, A. A., Ogundairo, J. A., & Danjibo, M. C. (2020). Coronavirus outbreak in Nigeria: Burden and socio-medical response during the first 100 days. *International Journal of Infectious Diseases*. https://doi.org/10.1016/j.ijid.2020.06.067

Arza, P. A., Ilham, D., & Hermaiyan, L. (2020). Pengaruh Pemberian Putih Telur Terhadap Kadar Kolesterol Total Pasien TB PARU. *Jurnal Kesehatan*. https://doi.org/10.35730/jk.v11i2.497

Asif, D., Bibi, M., Arif, M. S., & Mukheimer, A. (2023). Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyperparameter Optimization. *Algorithms*. https://doi.org/10.3390/a16060308

Awotunde, J. B., Folorunso, S. O., Ajagbe, S. A., Garg, J., & Ajamu, G. J. (2022). AiIoMT: IoMT-Based System-Enabled Artificial Intelligence for Enhanced Smart Healthcare Systems. In *Machine Learning for Critical Internet of Medical Things: Applications and Use Cases*. https://doi.org/10.1007/978-3-030-80928-7_10

Batteux, E., Mills, F., Jones, L. F., Symons, C., & Weston, D. (2022). The Effectiveness of Interventions for Increasing COVID-19 Vaccine Uptake: A Systematic Review. In *Vaccines*. https://doi.org/10.3390/vaccines10030386

Burnett, S. M., Mbonye, M. K., Naikoba, S., Stella, Z. M., Kinoti, S. N., Ronald, A., Rubashembusya, T., Willis, K. S., Colebunders, R., Manabe, Y. C., & Weaver, M. R. (2015). Effect of educational outreach timing and duration on facility performance for infectious disease care in Uganda: A trial with pre-post and cluster randomized controlled components. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0136966

Chen, T., Wang, Y., Chen, H., Marder, K., & Zeng, D. (2014). Targeted Local Support Vector Machine for Age-Dependent Classification. *Journal of the American Statistical Association*.

https://doi.org/10.1080/01621459.2014.881743

Dhanwanth, B., Vivek, B., Abirami, M., Waseem, S. M., & Manikantaa, C. (2023). Forecasting Chronic Kidney Disease Using Ensemble Machine Learning Technique. *International Journal on Recent and Innovation Trends in Computing and Communication*. https://doi.org/10.17762/ijritcc.v11i5s.7035

Dixon, S., Keshavamurthy, R., Farber, D. H., Stevens, A., Pazdernik, K. T., & Charles, L. E. (2022). A Comparison of Infectious Disease Forecasting Methods across Locations, Diseases, and Time. *Pathogens*. https://doi.org/10.3390/pathogens11020185

Huang, M., Li, Z., & Zhu, H. (2022). Recent Advances of Graphene and Related Materials in Artificial Intelligence. *Advanced Intelligent Systems*. https://doi.org/10.1002/aisy.202200077

Ibrahim, A., Abdelsalam, H., & Taj-Eddin, I. (2023). Software Defects Prediction At Method Level Using Ensemble Learning Techniques. *International Journal of Intelligent Computing and Information Sciences*. https://doi.org/10.21608/ijicis.2023.189934.1251

Jamshidi, M. B., Roshani, S., Talla, J., Lalbakhsh, A., Peroutka, Z., Roshani, S., Parandin, F., Malek, Z., Daneshfar, F., Niazkar, H. R., Lotfi, S., Sabet, A., Dehghani, M., Hadjilooei, F., Sharifi-Atashgah, M. S., & Lalbakhsh, P. (2022). A Review of the Potential of Artificial Intelligence Approaches to Forecasting COVID-19 Spreading. In *AI (Switzerland)*. https://doi.org/10.3390/ai3020028

King, T. E., Bradford, W. Z., Castro-Bernardini, S., Fagan, E. A., Glaspole, I., Glassberg, M. K., Gorina, E., Hopkins, P. M., Kardatzke, D., Lancaster, L., Lederer, D. J., Nathan, S. D., Pereira, C. A., Sahn, S. A., Sussman, R., Swigris, J. J., & Noble, P. W. (2014). A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa1402582

Kong, W., Zhu, J., Bi, S., Huang, L., Wu, P., & Zhu, S. J. (2023). Adaptive best subset selection algorithm and genetic algorithm aided ensemble learning method identified a robust severity score of COVID-19 patients. In *iMeta*. https://doi.org/10.1002/imt2.126

Kriss, J. L., Frew, P. M., Cortes, M., Malik, F. A., Chamberlain, A. T., Seib, K., Flowers, L., Ault, K. A., Howards, P. P., Orenstein, W. A., & Omer, S. B. (2017). Evaluation of two vaccine education interventions to improve pertussis vaccination among pregnant African American women: A randomized controlled trial. *Vaccine*. https://doi.org/10.1016/j.vaccine.2017.01.037

Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., & Cheng, L. (2023). A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. *Water (Switzerland)*. https://doi.org/10.3390/w15071265

Matloob, F., Ghazal, T. M., Taleb, N., Aftab, S., Ahmad, M., Khan, M. A., Abbas, S., & Soomro, T. R. (2021). Software defect prediction using ensemble learning: A systematic literature review. In *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3095559

Mbonye, M. K., Burnett, S. M., Naikoba, S., Ronald, A., Colebunders, R., Van Geertruyden, J. P., & Weaver, M. R. (2016). Effectiveness of educational outreach in infectious diseases management: A cluster randomized trial in Uganda. *BMC Public Health*. https://doi.org/10.1186/s12889-016-3375-4

Muthulakshmi, P., Parveen, M., & Rajeswari, P. (2023). Prediction of Heart Disease using Ensemble Learning. *Indian Journal Of Science And Technology*. https://doi.org/10.17485/ijst/v16i20.2279

OLIVEIRA, J. S., MACHADO, K. C., & FREITAS, R. M. (2014). Natural Products Applied A Neglected Diseases: Technological Forecasting. *Revista Gestão, Inovação e Tecnologias*. https://doi.org/10.7198/s2237-072220140002001

On, N. I., Boongoen, T., & Kongkotchawan, N. (2014). A new link-based method to ensemble clustering and cancer microarray data analysis. *International Journal of Collaborative Intelligence*. https://doi.org/10.1504/ijci.2014.064842

Peeri, N. C., Shrestha, N., Siddikur Rahman, M., Zaki, R., Tan, Z., Bibi, S., Baghbanzadeh, M., Aghamohammadi, N., Zhang, W., & Haque, U. (2021). The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? In *International Journal of Epidemiology*. https://doi.org/10.1093/IJE/DYAA033

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]. In *IEEE Computational Intelligence Magazine*. https://doi.org/10.1109/MCI.2015.2471235

Sarker, S., Jamal, L., Ahmed, S. F., & Irtisam, N. (2021). Robotics and artificial intelligence in healthcare during COVID-19 pandemic: A systematic review. In *Robotics and Autonomous Systems*. https://doi.org/10.1016/j.robot.2021.103902

Sharifi, A., & Khavarian-Garmsir, A. R. (2020). The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. In *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2020.142391

Singh, B., Vijayvargiya, A., & Kumar, R. (2022). Kinematic Modeling for Biped Robot Gait Trajectory Using Machine Learning Techniques. *Journal of Bionic Engineering*. https://doi.org/10.1007/s42235-021-00142-4

Solonen, A., Bibov, A., Bardsley, J. M., & Haario, H. (2014). Optimization-based sampling in ensemble Kalman filtering. *International Journal for Uncertainty Quantification*. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2014007658

Vammi, V., Lin, T. L., & Song, G. (2014). Enhancing the quality of protein conformation ensembles with relative populations. *Journal of Biomolecular NMR*. https://doi.org/10.1007/s10858-014-9818-2

Wu, H., & Levinson, D. (2021). The ensemble approach to forecasting: A review and synthesis. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2021.103357

Ye, J. (2020). The role of health technology and informatics in a global public health emergency: Practices and implications from the COVID-19 pandemic. In *JMIR Medical Informatics*. https://doi.org/10.2196/19866

Zheng, R., & Liu, G. (2023). Application of machine learning in clinical predictive models for infectious diseases: a review. *Chinese Journal of Schistosomiasis Control*. https://doi.org/10.16250/j.32.1374.2023084

Zhu, J., Zhang, A., & Zheng, H. (2023). Research on Predictive Model Based on Ensemble Learning. *Highlights in Science, Engineering and Technology*. https://doi.org/10.54097/hset.v57i.10023