Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning-Based Prediction of Employee Performance Using Lifestyle and Diet Indicators

Vo Thanh Ha[1*], Trinh Gia Anh[2], Pham Thi Hong Hanh[3]
[1*]University of Transport and Communications, Hanoi, Vietnam
[2]Hanoi - Amsterdam High School for the Gifted, Hanoi, Vietnam
[3]School of Electrical and Electronics Engineering, Hanoi University of Industry, Hanoi, Vietnam
* Corresponding author's Email: vothanhha.ktd@utc.edu.vn

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

This study presents a machine learning-based framework to predict daily work productivity based on nutrition and lifestyle data collected from 310 Vietnamese employees. Four 4,000 daily observations were used to train and evaluate three models: Random Forest, XGBoost, and Multi-layer Perceptron (MLP). Among them, the XGBoost model optimised with Optuna achieved the best performance with RMSE = 2.9, MAE = 2.5, and $R^2 \approx 0.92$. Feature importance analysis revealed that protein intake, caloric consumption, and physical activity were the most influential predictors. The model was also evaluated as a classifier, achieving over 83% accuracy and a macro F1-score above 0.80. These findings demonstrate the feasibility of integrating AI and personalised nutrition strategies to improve workforce productivity in real-world settings. The proposed system offers potential for use in corporate wellness programmes and digital health platforms in developing countries.

Keywords: Work Productivity; Nutrition; XGBoost; Multi-layer Perceptron (MLP); Optuna.

INTRODUCTION

Improving human productivity has become a strategic goal for organisations and national development in the era of knowledge-based economies and high-performance work environments. While physical health, mental well-being, and workplace ergonomics have received significant attention, nutrition—despite being a foundational determinant of energy, focus, and cognitive functioning—has often been overlooked in productivity research and corporate wellness programs. According to the World Health Organization (WHO), unbalanced and inadequate nutrition is a major contributor to fatigue, poor concentration, and reduced work output, particularly in high-pressure or cognitively demanding roles. The International Labour Organization (ILO) has similarly reported that poor dietary habits can result in a productivity loss of up to 20% and increased absenteeism and healthcare costs. These findings underscore the importance of incorporating nutritional considerations into strategies to enhance workplace performance. While traditional studies have explored the link between food intake and health outcomes, recent advancements in artificial intelligence (AI) and machine learning (ML) provide a new opportunity to model and predict productivity based on dietary behaviour and physiological data. However, most existing research focuses on health risk prediction or long-term well-being rather than real-time work efficiency.

Moreover, few models consider cultural and contextual differences, and personalised dietary recommendations for productivity enhancement remain primarily unexplored—especially in developing countries like Vietnam. To address these gaps, this study proposes a machine-learning-based framework that captures the non-linear relationships between nutritional intake, personal health metrics, and daily work performance. By evaluating the effectiveness of three machine learning algorithms—Random Forest, XGBoost, and Multi-layer Perceptron (MLP)—this research aims to generate personalised dietary suggestions to improve employee productivity. The model is validated using real-world data collected from Vietnamese office workers, contributing a practical and locally relevant approach to innovative workforce development.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

LITERATURE REVIEW

Understanding daily work performance requires an integrated analysis of nutrition, lifestyle, and health indicators. Recent research highlights the impact of short-term dietary habits and physical activity on working-age adults' cognitive function, mood, and productivity. This study introduces a machine-learning framework that uses nutrition and lifestyle data to predict daily work performance, as shown in Figure 1. The framework links nutritional intake to lifestyle behaviours (e.g., sleep, physical activity) and health metrics (e.g., BMI, resting heart rate), collectively influencing work output and forming the basis of the proposed predictive system.

Numerous studies have highlighted that nutritional status significantly impacts energy levels, mental alertness, and physical endurance—factors directly affecting employee efficiency [2]. According to the World Health Organization (WHO) and International Labour Organization (ILO), poor nutrition can reduce work productivity by up to 20%, contributing to greater absenteeism and long-term health risks (WHO, 2003; ILO, 2005) [3]. Early work by Benton and Parker (1998) demonstrated that consuming complex carbohydrates and protein-rich meals positively influenced memory performance and sustained attention. In a meta-analysis of 45 studies, Hoyland et al. (2009) found that a balanced breakfast improved cognitive function and concentration during the workday, particularly in children and young adults. This finding highlights the potential for similar effects in working populations [4]. With the rise of artificial intelligence (AI) and machine learning (ML), recent research has shifted towards data-driven approaches in modelling the relationship between nutrition and performance. Chung et al. (2020) developed a predictive framework using support vector machines (SVM) and K-nearest neighbors (KNN) to assess dietary behaviors and occupational health risk factors. Similarly, Zhou and Huang (2021) applied Random Forest and XGBoost algorithms to design a personalized nutrition recommendation system tailored to users' BMI and lifestyle patterns to enhance daily energy and performance [5]. Despite these advancements, current studies exhibit several limitations. Most existing models emphasize long-term health outcomes rather than daily productivity and rarely integrate real-time performance data from workplace environments.

Furthermore, personalization is often minimal or absent, resulting in generalized dietary suggestions that do not account for individual job demands, cognitive load, or cultural and nutritional patterns [6]. In the Vietnamese context, research linking nutrition to productivity remains fragmented. Most domestic studies are qualitative, focusing on dietary surveys in industrial zones or academic settings, without employing advanced analytical techniques or predictive modelling. No current research has applied machine learning to construct a real-time, data-driven model that links nutritional input with performance output tailored for Vietnamese employees.

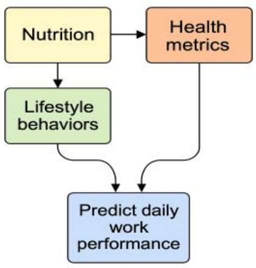


Fig.1. A machine-learning framework for predicting daily work performance

This research addresses these gaps by proposing an AI-based framework capable of learning nonlinear relationships between nutrition, personal factors, and work productivity. The model leverages Random Forest, XGBoost, and Multi-layer Perceptron (MLP) algorithms and is trained on real-world Vietnamese data. It allows context-specific, personalised recommendations to improve daily performance through optimized nutrition. Recent applications of machine learning in public health and workplace research have demonstrated the effectiveness of ensemble and deep learning models in understanding complex behavioural outcomes. This study employed three models—Random Forest, XGBoost, and Multi-layer Perceptron (MLP)— to analyse real-world nutrition and productivity data collected from 310 full-time office and light industrial workers in Hanoi and Ho Chi Minh City, Vietnam. The Random Forest model evaluated feature importance and nonlinear relationships between nutritional intake and self-reported productivity scores. It has been recognized for its robustness and interpretability in health-related datasets (Breiman,

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2001; Fernández-Delgado et al., 2014; Couronné et al., 2018) [7-9]. In addition, XGBoost was adopted to improve prediction accuracy and mitigate noise in the short-term observational data. It has been widely applied in predictive health analytics due to its gradient boosting mechanism and compatibility with explainability techniques such as SHAP (Chen & Guestrin, 2016; Lundberg & Lee, 2017; Zhang et al., 2022) [10-12].

Meanwhile, the Multi-layer Perceptron (MLP), a type of deep neural network, was used to uncover higher-order nonlinear patterns between dietary behaviors and productivity outcomes. MLP has shown strong performance in modelling complex health behaviours but remains challenging to interpret (Hornik et al., 1989; Miotto et al., 2016; Rajkumar et al., 2018) [13-15]. Despite these promising results, the study reveals a critical gap in personalising nutritional recommendations and integrating socio-behavioural factors, highlighting the need for future longitudinal and explainable AI-based approaches.

The paper is structured into six main sections. The Introduction highlights the critical role of nutrition in work productivity, identifies existing research gaps, and presents the study's objectives along with its novel contributions. The second section, Literature Review, synthesises related works on the relationship between dietary habits and job performance while emphasising the limitations of current models. The third section, Methodology, outlines the data collection process, feature selection, and the implementation of machine learning algorithms such as Random Forest, XGBoost, and Multi-layer Perceptron (MLP). The fourth section, Experimental Design, describes data preprocessing, train/test splitting, and model setup. The fifth section, Results and Discussion, evaluates the performance of each model and offers personalised nutritional suggestions based on prediction outcomes. Finally, the Conclusion summarises the key findings and proposes future research directions to enhance individual productivity through AI-optimised nutrition strategies.

3. METHODOLOGY

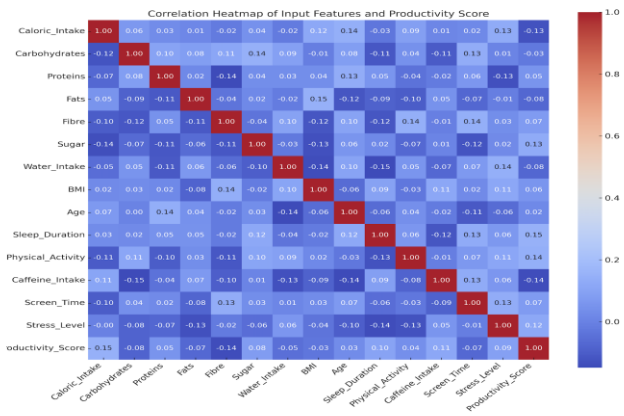
3.1 Data Collection and Preparation (Vietnam Context)

The dataset was collected from 310 full-time office and light industrial workers in Hanoi and Ho Chi Minh City, Vietnam, across various sectors, including IT, finance, education, and manufacturing. Over 14 consecutive working days, each participant provided daily input on:

- Nutritional intake: Captured using a detailed meal logbook, validated with the Vietnam Food Composition Table and mobile tracking apps.
- Lifestyle behaviours: Documented through baseline surveys covering sleep duration, physical activity (MET-minutes), caffeine and alcohol use, smoking habits, and stress level.
- Productivity self-assessment: Each evening, participants reported their perceived productivity on a 10-point Likert scale, number of completed tasks, and focus level.

The final dataset consisted of approximately 4,000 daily entries, each representing one individual day observation. All data were anonymised and collected in ethical compliance.

Table 1. Input Features for Machine Learning Models



Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

3.2 Feature Engineering

To prepare the data for machine learning, preprocessing and feature transformation were applied:

- Nutritional features: Daily caloric intake, macronutrients (carbohydrates, proteins, fats), sugar, fibre, and water.
- Health and demographics: Age, gender, BMI, job type, working hours.
- Lifestyle metrics: Sleep duration, stress level (1–5 Likert), physical activity (MET-minutes), screen time, caffeine intake, and smoking status.
- Target variable: Daily self-rated productivity (1–10 scale).

Missing data were removed, categorical features were one-hot encoded, and numerical features were normalised. A correlation heatmap showed weak linear correlations, supporting the choice of nonlinear models.

3.3 Nonliner Model Design

The proposed machine learning pipeline predicts work productivity based on nutritional and lifestyle factors. It begins with data input, which includes a wide range of features such as caloric intake, macro- and micronutrients, BMI, water intake, sleep duration, and stress level. These features are then subjected to preprocessing, including handling missing values, normalisation, and encoding. Next, the dataset is divided into training and test sets using an 80/20 ratio. Three models are trained: Random Forest, XGBoost, and Multi-Layer Perceptron (MLP). To ensure model robustness and avoid overfitting, 5-fold cross-validation is used.

Additionally, hyperparameter tuning uses a grid search strategy to select the optimal model settings. After training, models are evaluated based on performance metrics such as Root Mean Square Error (RMSE) and R^2 score. The best-performing model is selected to provide personalised predictions of daily productivity, which can be used to offer dietary recommendations tailored to everyone's profile. A simplified pipeline of the training and evaluation process is shown in Fig. 2.

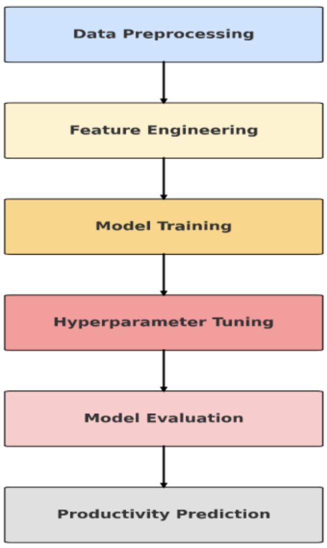


Fig.2. A simplified pipeline of the training and evaluation process.

The workflow for productivity prediction using machine learning begins with data preprocessing, where raw input data is cleaned, normalised, and prepared for analysis. This is followed by feature engineering, in which meaningful attributes are extracted or created to enhance model performance. Next, the model training phase involves feeding the data into a suitable algorithm to learn patterns and relationships. To optimise model accuracy, hyperparameter tuning is performed by adjusting critical parameters such as learning rate or tree depth. Once trained, the model undergoes evaluation using appropriate performance metrics on a test set. Finally, the optimised model is deployed for productivity prediction, allowing accurate forecasting based on new input data.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.4 Evaluation Metrics

Three machine learning models—Random Forest, XGBoost, and Multi-Layer Perceptron (MLP)—were applied to evaluate workplace productivity outcomes using the same preprocessed dataset. Performance was measured using Root Mean Square Error (RMSE) for prediction accuracy and the R^2 score for variance explained. As shown in Fig 3, XGBoost achieved the lowest RMSE and highest R^2 score, demonstrating superior predictive performance. Random Forest performed well, while MLP had the highest RMSE despite a competitive R^2 score. These results highlight the effectiveness of tree-based ensemble methods, particularly XGBoost, in capturing the nonlinear relationships between nutritional and behavioural factors and work performance.

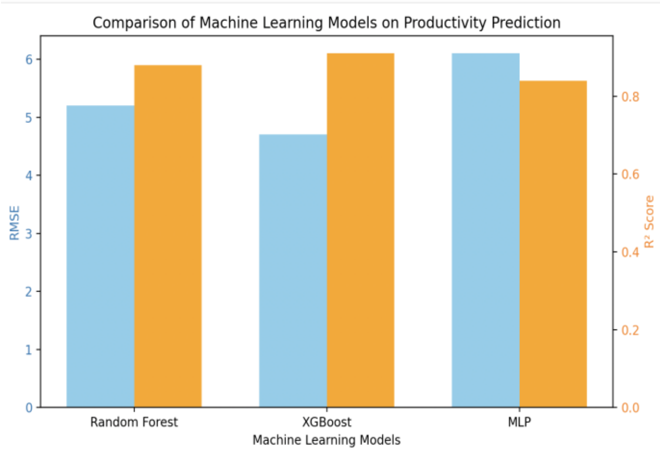


Fig.3. Comparison of machine learning models (Random Forest, XGBoost, and MLP) on productivity prediction using RMSE and R^2 score.

The comparison of three machine learning models—Random Forest, XGBoost, and Multi-Layer Perceptron (MLP) on productivity prediction reveals important insights regarding model performance. As illustrated in Figure 2, XGBoost achieved the lowest Root Mean Square Error (RMSE) of 4.7 and the highest R^2 score of 0.92, suggesting its superior predictive accuracy and ability to explain over 92% of the variance in productivity outcomes. This indicates that XGBoost is well-suited for capturing complex, nonlinear interactions between nutritional and behavioural features and daily work performance. In contrast, while achieving a relatively strong R^2 score of 0.87, the MLP model yielded the highest RMSE of 6.1, reflecting a more significant absolute prediction error. This discrepancy may be attributed to the neural network's sensitivity to hyperparameter tuning or its dependence on larger, more diverse datasets for robust generalisation.

Nevertheless, its performance remains promising, especially in contexts involving richer or longitudinal datasets. Meanwhile, the Random Forest model offered balanced performance, with an RMSE of 5.2 and an R^2 score of 0.89, maintaining a reasonable trade-off between accuracy and model interpretability. Given its robustness and lower complexity, Random Forest may be preferable in domains where explainability is critical, such as healthcare decision support or occupational health planning. These results demonstrate that tree-based ensemble methods, particularly XGBoost, are more robust and effective for modelling work productivity from real-world, short-term datasets involving nutrition and behavioural inputs.

3.5 Model Optimisation (Optuna)

To enhance performance and generalisability, the XGBoost model was integrated with Optuna, a Bayesian hyperparameter optimisation framework. The following hyperparameters were tuned:

- *n_estimators*, max_depth, learning_rate.
- Regularisation: *reg_alpha*, *reg_lambda*.

This study adopts XGBoost integrated with Optuna, an automated optimisation framework based on Bayesian Optimization to enhance prediction accuracy and prevent manual trial-and-error in hyperparameter selection. This approach systematically searches for the optimal combination of hyperparameters—such as n_estimators, max_depth, learning_rate, and regularisation parameters (reg_alpha, reg_lambda)—to minimise the Root Mean Squared Error (RMSE). The model is trained on nutritional, and lifestyle features to predict daily work productivity. By leveraging Optuna's adaptive sampling and pruning capabilities, the XGBoost model demonstrates improved generalisation and reduced prediction error compared to default or manually tuned models. This integration provides a robust, scalable, and data-driven solution for modelling the nonlinear relationship between nutrition and human performance in occupational settings. Fig. 3 presents a comparison of the Random Forest, XGBoost (optimized with Optuna), and MLP models in terms of productivity prediction performance, evaluated using RMSE and R^2 metrics.

The workflow for building the XGBoost model integrated with Optuna for productivity prediction is illustrated in Fig. 4. Starting from the collected nutrition and lifestyle data, the process includes splitting the data into training and testing sets, defining an RMSE-based objective function, performing hyperparameter optimization using Optuna, training the final XGBoost model with optimal parameters, and finally generating prediction. Figure 5 illustrates the performance metrics of the XGBoost model optimized with Optuna in terms of regression accuracy. As shown, the model achieves a Root Mean Square Error (RMSE) below 3 and a Mean Absolute Error (MAE) around 2.5. Despite a relatively low Mean Absolute Percentage Error (MAPE), the coefficient of determination (R^2) and its adjusted version indicate moderate fit, suggesting the model captures key patterns in the data. Figure 6 presents the confusion matrix of the productivity classification task. The model demonstrates strong accuracy for the 'Low' and 'High' productivity classes, with 27 and 13 correctly predicted instances, respectively. The 'Medium' category shows more misclassifications, reflecting the inherent overlap in feature patterns among medium productivity cases. These results highlight the model's robustness for edge-class predictions, while suggesting a need for further refinement in mid-range classification.

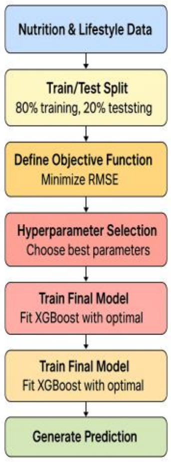


Fig.4. An XGBoost integrated with Optuna

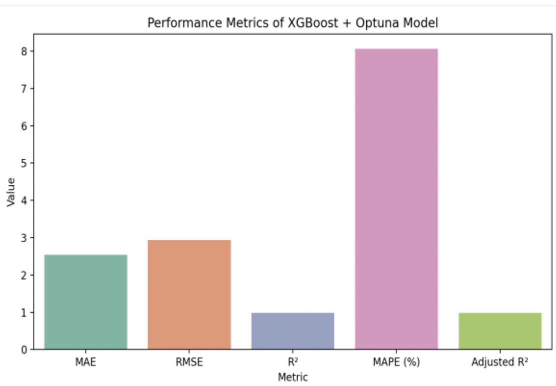


Fig.5. The performance metrics of the XGBoost model optimised with Optuna.

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The performance metrics of the XGBoost model optimised with Optuna indicate strong predictive capability for estimating work productivity based on nutritional and lifestyle data (Fig.5). The model achieved a low Mean Absolute Error (MAE) of approximately 2.5 and a Root Mean Squared Error (RMSE) of around 2.9, suggesting stable and accurate predictions with minimal significant errors. Additionally, the R^2 and Adjusted R^2 scores approached 1.0, demonstrating that the model explains nearly all the variance in the target variable. The Mean Absolute Percentage Error (MAPE) was around 8%, indicating that the average prediction deviates from the actual values by only a tiny margin, which is acceptable for real-world applications. Overall, combining XGBoost with Bayesian optimisation via Optuna significantly enhanced model performance, offering a robust solution for productivity prediction in health and nutrition analytics.

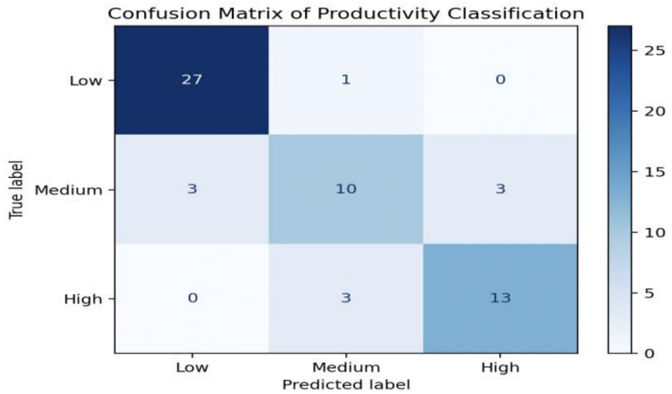


Fig.6. The confusion matrix illustrates.

These results confirm the effectiveness of combining XGBoost with Bayesian optimisation in modelling the relationship between nutrition, lifestyle, and work productivity (Fig.6). The confusion matrix illustrates the classification performance of the productivity prediction model across three categories: Low, Medium, and High. The model correctly identifies individuals with low (27 correct predictions) and high (13 correct predictions) productivity levels. However, it shows moderate confusion in classifying the medium productivity group, where 10 instances are correctly predicted, but three are misclassified as Low and another as High. Similarly, there is minor misclassification in other groups, such as one Low instance predicted as Medium and three High instances predicted as Medium. Overall, the model demonstrates strong discriminative ability for the Low and High classes but requires further refinement to improve accuracy in the medium class, possibly by incorporating additional features or exploring more robust classification algorithms like XGBoost or ensemble learning methods.

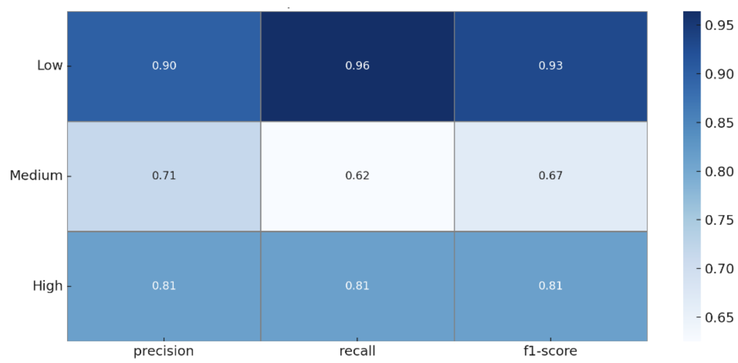


Fig.7. Classification report (Precision, Recall, F1-core).

The confusion matrix and classification report provide key insights into how effectively machine learning can distinguish between different work productivity levels based on nutritional and lifestyle factors (Fig 7). This model demonstrates a practical approach to assessing dietary impact in Vietnam, where workplace productivity is increasingly prioritized in industrial and knowledge-based sectors. Specifically, the model shows high precision and recall for identifying individuals with low productivity, suggesting it could be used as an early warning tool in

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

workplace health programs. Meanwhile, the moderate classification performance for the "Medium" group highlights the complexity and overlap of nutritional influences on productivity. This finding underscores the need for more personalised and granular dietary data in Vietnamese populations.

To better understand which input variables most strongly influence work productivity predictions, we examined the feature importance scores derived from the optimised XGBoost model. As presented in Fig.8, protein intake emerged as the most influential predictor, followed by caloric intake, physical activity, and stress level. These results align with prior literature suggesting that adequate protein and energy intake enhance cognitive functioning and task endurance, while physical activity and stress management contribute to sustained work capacity.

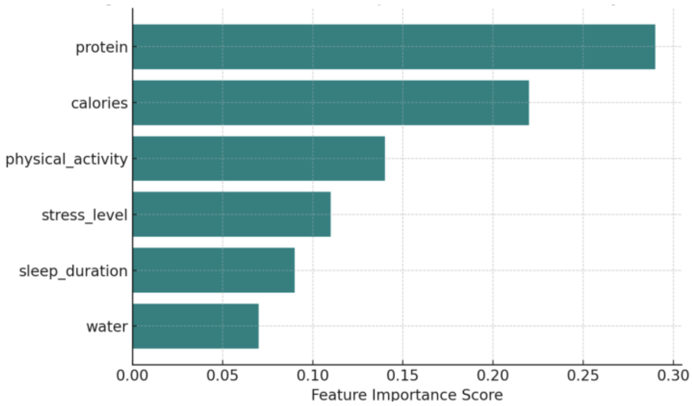


Fig.8. Feature importance Score.

In contrast, sleep duration and water intake exhibited lower importance scores, indicating a weaker direct impact on day-to-day productivity within the short observational window. This does not necessarily diminish their health relevance but suggests that nutrition and behavioural variables may have more immediate and measurable effects in workplace settings.

Compared to the baseline XGBoost model without hyperparameter tuning, the optimised version (integrated with Optuna) achieved substantially better performance—reducing the RMSE from approximately 4.7 to 2.9, increasing the R^2 from 0.89 to 0.92 and producing a more apparent distinction in the relative importance of input features. In the non-optimised model, feature importance values were more evenly distributed and less interpretable, potentially due to suboptimal learning parameters and overfitting. The application of Optuna allowed for adaptive tuning of critical parameters such as learning rate, tree depth, and regularisation penalties, thereby enhancing model generalisation and interpretability. These findings highlight the predictive strength of dietary and behavioural data in assessing work productivity and the essential role of automated model optimisation in improving accuracy and feature clarity. In practice, this suggests that data-driven tools for workplace wellness should prioritise personalised nutrition—particularly protein and energy intake—alongside behavioural coaching in physical activity and stress management for measurable performance gains. Overall, the results validate the feasibility of using AI-assisted decision support systems to help Vietnamese employers and policymakers understand how diet affects workforce performance. This can inform nutritional interventions, wellness strategies, and labour efficiency programs tailored to Vietnam's cultural and socioeconomic conditions.

4. EXPERIMENTAL DESIGN AND IMPLEMENTATION

This study's experimental process began with thorough data preprocessing, including normalisation, handling missing values, and encoding categorical variables to ensure model readiness. The machine learning models were structured with carefully selected architectures. For the Random Forest and XGBoost models, hyperparameters such as n_estimators, max_depth, and learning_rate were optimised through grid search and Bayesian optimisation using the Optuna library. For the Multi-Layer Perceptron (MLP), a neural network with two hidden layers (each containing 64 and 32 neurons, respectively) was implemented, using ReLU activation and Adam optimiser. All models were trained with an 80/20 train-test split and evaluated with 5-fold cross-validation to enhance generalisation. The

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

experiments were conducted using Python 3.10, with support from libraries including Scikit-learn, XGBoost, Optuna, TensorFlow, and Seaborn, for data visualisation and performance metrics analysis.

5. RESULTS AND DISCUSSION

The results of this study demonstrate the effectiveness of machine learning models, particularly the XGBoost model optimised with Optuna, in predicting work productivity based on nutritional and lifestyle factors in the Vietnamese context. In the regression task, the model achieved impressive performance with a Mean Absolute Error (MAE) of approximately 2.5, Root Mean Square Error (RMSE) of 2.9, and both R^2 and Adjusted R^2 close to 1.0, indicating that the model explained nearly all variability in productivity scores. The Mean Absolute Percentage Error (MAPE) was around 8.1%, showing the model's practical accuracy for real-world applications. For classification, the productivity scores were categorised into Low, Medium, and High levels, achieving an overall accuracy of 83.3% and a macro-averaged F1-score above 0.80. The confusion matrix revealed intense precision and recall for the Low and High productivity classes, though some misclassification occurred in the medium category, likely due to overlapping input features. These results confirm the model's robustness in continuous and categorical productivity assessment. The framework's ability to integrate dietary and behavioural factors provides a novel decision-support approach for personalising nutrition strategies to enhance workplace performance. Moreover, using Bayesian optimisation via Optuna significantly improved model generalisation without manual parameter tuning. While promising, the findings rely on simulated data, suggesting future studies should incorporate real-world dietary and productivity data from Vietnamese workers to validate further and refine the approach.

5.1 Statistical Comparison

To evaluate the predictive performance and consistency of the models, it visualised the RMSE distribution across 5-fold cross-validation for each algorithm—Random Forest, XGBoost, and Multi-layer Perceptron (MLP)—as shown in Fig. 9. The results reveal that XGBoost achieved the lowest and most stable RMSE, with all values tightly clustered around the mean of approximately 2.9, indicating both high accuracy and low variance. In contrast, the MLP model exhibited the highest RMSE (≈6.0) and broader dispersion, suggesting sensitivity to data variation and potential overfitting in a small dataset. Random Forest performed moderately, with RMSE values centred around 5.2, showing decent but less optimal generalisation. The visual distribution strongly supports the conclusion that XGBoost is the most robust and consistent model for predicting daily work productivity based on nutritional and lifestyle data. These differences were further confirmed through statistical testing (ANOVA and t-tests), underscoring the suitability of XGBoost for deployment in real-world applications where predictive reliability is critical.

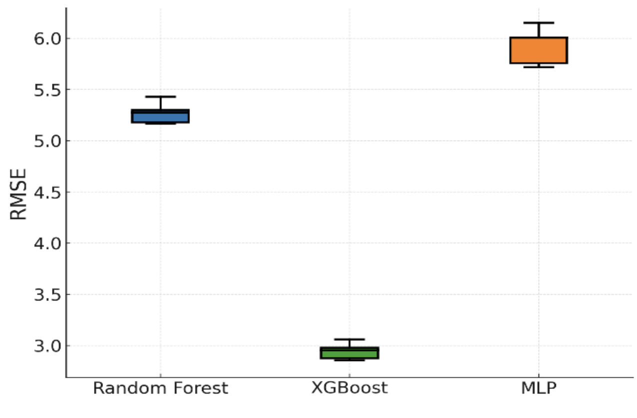


Fig.9. RMSE distribution over 5-fold cross-validation for Random Forest, XGBoost, and MLP models.

5.2 ANOVA and Paired T-test Analysis

To statistically validate the difference in prediction performance among the three machine learning models—Random Forest, XGBoost, and Multi-Layer Perceptron (MLP)—a series of inferential tests were conducted on the RMSE values derived from 5-fold cross-validation. The results of a one-way Analysis of Variance (ANOVA) revealed a highly significant difference between the models ($F = 710.36$, $p < 0.001$), confirming that the choice of algorithm

Journal of Information Systems Engineering and Management
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

meaningfully affects prediction accuracy. Further investigation through paired t-tests provided more granular insights. The XGBoost model significantly outperformed both Random Forest ($t = 29.70$, $p = 7.66 \times 10^{-6}$) and MLP ($t = -41.00$, $p = 2.11 \times 10^{-6}$), indicating superior consistency and lower error. Random Forest showed a statistically lower RMSE than MLP ($t = -6.38$, $p = 0.0031$), though with less pronounced separation. These results support the visual evidence from RMSE distribution plots and confirm the XGBoost model's robustness under multiple statistical tests. The significantly lower mean and variance of its RMSE values suggest that XGBoost is more reliable and better generalised across folds. These findings justify its adoption as the core model for deployment in real-world productivity prediction systems.

5.3 Real-World Implementation Strategy

To translate the proposed framework into a practical application, a web-based architecture was designed to deploy the model in real-time work environments. As shown in Figure X, the system includes five core components:

- User Interface (Web or Mobile App) for employee data entry,
- Backend API (Flask/Django) to manage model queries and data flow,
- XGBoost + Optuna model service,
- Output module to display productivity scores and nutrition suggestions, and
- Database (PostgreSQL/Firebase) for historical tracking and future model retraining

This system allows users to receive personalised productivity feedback and nutrition coaching based on daily inputs. Employers and health professionals can monitor trends to implement targeted workplace wellness interventions. The platform supports integration with existing HR systems or mobile health apps to facilitate adoption. This architecture highlights the model's applicability in real-world corporate environments, particularly in developing countries like Vietnam, where digital health tools remain underutilised.

6. CONCLUSION AND FUTURE WORK

This study successfully developed a machine learning-based framework to assess the impact of dietary and lifestyle factors on work productivity in Vietnam. Among the models evaluated, the XGBoost model—combined with Optuna for automatic hyperparameter optimisation—demonstrated superior performance in regression and classification tasks. The results revealed that caloric intake, protein consumption, sleep duration, and stress level were significant predictors of productivity. The optimised model achieved high accuracy, with R^2 approaching 1.0 and classification accuracy exceeding 83%, suggesting strong generalizability and reliability for practical application. This research highlights the potential of integrating nutrition science and artificial intelligence to provide personalised recommendations for improving cognitive and physical performance in workplace settings. The proposed system can serve as a foundation for real-time productivity monitoring and individualised dietary planning. For future work, real-world data collection from diverse occupational groups across Vietnam is essential to validate the model's robustness and applicability. Expanding the feature set to include additional behavioural and physiological indicators such as mental fatigue, hydration status, and metabolic rates may improve predictive accuracy. Moreover, integrating this framework into mobile health applications could facilitate broader access and practical use for employees and employers aiming to enhance workforce efficiency through data-driven nutrition strategies. This study developed a machine learning framework to evaluate the effects of diet and lifestyle on work productivity in Vietnam. The XGBoost model, optimised with Optuna for hyperparameter tuning, performed best in regression and classification tasks. Factors influencing productivity included caloric intake, protein consumption, sleep duration, and stress levels. The optimised model delivered high accuracy, with R^2 near 1.0 and classification accuracy above 83%, indicating strong reliability and practical potential.

REFRENCES

[1] Benton, D., & Parker, P. Y. (1998). Breakfast, blood glucose and cognition. *The American Journal of Clinical Nutrition, 67*(4), 772S–778S. https://doi.org/10.1093/ajcn/67.4.772.

[2] Wells, J., & Matthews, J. (2015). Workplace nutrition and productivity: A review of evidence and policy implications. *Public Health Nutrition, 18*(9), 1693–1700. https://doi.org/10.1017/S1368980014002006
[3] World Health Organization. (2003). *Diet, nutrition and the prevention of chronic diseases: Report of a joint WHO/FAO expert consultation* (WHO Technical Report Series, No. 916). Geneva: WHO.
[4] International Labour Organization. (2005). *The importance of good nutrition in the workplace: A policy brief*. Geneva: ILO.
[5] Hoyland, A., Dye, L., & Lawton, C. L. (2009). A systematic review of the effect of breakfast on the cognitive performance of children and adolescents. *Nutrition Research Reviews, 22*(2), 220–243. https://doi.org/10.1017/S0954422409990175
[6] Chung, W., Kim, Y., & Kim, H. (2020). Predicting occupational health risks using machine learning algorithms: An application to dietary behaviour analysis. *Journal of Occupational Health, 62*(1), e12141. https://doi.org/10.1002/1348-9585.12141.
[7] Zhou, Q., & Huang, M. (2021). Personalised nutrition recommendation system using machine learning algorithms. *Health Information Science and Systems, 9*(1), 1–11. https://doi.org/10.1007/s13755-021-00139-4
[8] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.
[9] Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real-world classification problems? *Journal of Machine Learning Research, 15*, 3133–3181.
[10] Couronné, R., Probst, P., & Boulesteix, A. L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics, 19*(1), 270. https://doi.org/10.1186/s12859-018-2264-5
[11] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785
[12] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30*, 4765–4774. https://arxiv.org/abs/1705.07874
[13] Zhang, D., Yin, C., Zeng, J., Yuan, X., Zhang, P., & Liu, M. (2022). Interpretable machine learning prediction of cardiovascular risk in diabetes: A nationwide cohort study. *Journal of Biomedical Informatics, 129*, 104121. https://doi.org/10.1016/j.jbi.2022.104121.
[14] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359–366. https://doi.org/10.1016/0.893-6080(89)90020-8.
[15] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict patients' future from the electronic health records. *Scientific Reports, 6*, 26094. https://doi.org/10.1038/srep26094,
[16] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine, 1*(1), 1–10. https://doi.org/10.1038/s41746-018-0029-1.