Advancing Video Game Addiction Detection: A Stacking Ensemble Approach Utilizing Machine Learning and Deep Learning Models

Amina BOUHADJA[1,*,†], Abdelkrim BOURAMOUL[2]

[1]University of Skikda, LICUS Laboratory, Algeria
[2]University of Abdelhamid Mehri Constantine 2, MISC Laboratory, Algeria

Abstract

Video game addiction is a phenomenon characterized by excessive and compulsive dependence on video games. This addiction can lead to mental health problems and detrimental consequences in other areas of life. Questionnaires and interviews used to identify this dependence can be influenced by biases and errors, requiring active participation from the individuals involved. Moreover, these methods can be inefficient and require significant resources in terms of time and costs. To overcome these limitations, we have leveraged advances in machine learning and deep learning. Utilizing data on gaming habits, playtime durations, demographic data, and other relevant features, our model effectively identifies signs of addiction. The development process followed an incremental approach, beginning with the creation of various models, including decision trees, Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), Feed-Forward Neural Networks (FFNN), and random forests. These models were chosen for their popularity and widespread use in the field. Recognizing shortcomings in individual models, we transitioned to an ensemble technique known as "Stacking" to address overfitting issues and enhance overall performance. The final selection of the appropriate model aimed at reducing complexity. The stacking model demonstrated notable accuracy scores Accuracy: 0.95, Recall: 0.89, Precision: 0.87.

Keywords

Gaming disorder, Machine Learning, Ensemble Learning, Stacking Learning, Logistic Regression.

1. Introduction

Nowadays, video game disorder is a growing problem that raises within our modern society. With the rising popularity of video games and their easy accessibility through various platforms, young individuals find themselves struggling with excessive and compulsive addiction to video games, this addiction can lead to detrimental consequences for their mental health, academic performance, social relationships, and overall well-being. In this context, the preventive prediction of video game addiction aims to identify at-risk individuals before the addiction becomes problematic. The goal is to intervene early by providing tailored support, prevention strategies, and therapeutic practices. The utilization of Artificial Intelligence (AI) methodologies, such as machine learning and deep learning, has facilitated significant advancements, it is possible to analyze vast amounts of data related to gaming habits, online behaviors, and other factors to identify early signs of addiction. In our study, we developed several individual models, each trained on preprocessed data to predict video game addiction. However, as we conducted our experiments, we observed a phenomenon of overfitting. This occurs due to the limitations of the data we have, leading to poor generalization on new data. To overcome the overfitting issue, we opted for the ensemble learning method known as Stacking. This method allows us to combine the predictions of several individual models to obtain a more accurate final prediction. The use of Stacking enabled us to reduce overfitting and biases. Stacking relies on the use of a meta-model that learns by combining predictions from multiple individual models rather than relying solely on an individual prediction. The meta-model integrates predictions from each individual model and produces a more reliable final prediction by leveraging the strengths of each individual model. We were inspired

RIF 2024: The 13th Conference on Research in computIng at Feminine, May 20–21, 2024, Constantine, Algeria
amina.bouhadja@univ-skikda.dz (A. BOUHADJA); abdelkrim.bouramoul@univ-constantine2.dz (A. BOURAMOUL)
0009-0003-3408-8581 (A. BOUHADJA); 0000-0002-8628-6817 (A. BOURAMOUL)

by the architecture of Stacking ensemble learning proposed in our previous article, which utilized heterogeneous machine learning and deep learning architectures to predict subsequent substance use disorder based on initial addiction [10].
In order to enhance the originality of our current work, we enriched our Stacking-based architecture by adding a significant extension. This extension aims to select the most suitable model from the trained individual models for the given data. By using coefficients for selection, the selected model stood out as the most suitable for the dataset's characteristics. The decision to select this model is motivated by two factors: first, the increased complexity of the stacking architecture when dealing with large datasets, which results in longer computation times. Second, stacking models are often less interpretable compared to models with simpler architectures. Enriching the architecture allowed us to strike a balance between prediction accuracy and interpretability of the results, which is important for experts in the field of video game addiction to better understand the reasons for video game dependency.
The rest of the paper is organized as follows: Section 2 provides an overview of the most recent relevant research. Section 3 provides an overview of the employed methodologies and tools, as well as a detailed explanation of the proposed architectures. Section 4 presents the empirical investigation and an elaborate analysis of the findings. Section 7 ends with conclusion and outlines the intentions for future investigations.

2. Related Work

In this section, we present research conducted in the same field as our article, focusing on the application of machine learning and deep learning techniques to study video game addiction. Specifically, the results of these studies and the techniques used will be discussed, along with an attempt to identify their limitations.

The contribution proposed by Han, Xu, et al. [1] centered on the use of radiomics and machine learning techniques to distinguish brain differences between individuals with gaming addiction disorders and healthy individuals. The authors employed a promising radiomics-based method to extract features from brain MRI images. Subsequently, machine learning was utilized to train a classification model capable of identifying both groups, namely healthy individuals and those with gaming disorders. The results revealed that the Random Forest model could distinguish between the two groups with an accuracy of 73%. However, it is important to note that the study was conducted on a small sample size, potentially impacting the model's generalizability. In conclusion, while the use of radiomics and machine learning to identify brain differences in gaming disorder subjects is promising, larger and well-designed studies are needed to validate these results.

In a similar context, another study conducted in 2019 [2] used machine learning to identify individuals with a desire to play video games based on multimodal physiological signals, including skin conductance and eye movements. Data were collected using sensors such as electrodes to measure skin conductance, and participants also responded to standardized questionnaires assessing their desire to play video games. Results showed that this method could detect the desire to play video games with an accuracy of up to 80% based on multimodal physiological signals. However, the study participants were examined in a laboratory rather than their natural environment, potentially limiting the validity of the results, as laboratory experiences may differ from those in individuals' natural environments. Therefore, machine learning models may not generalize, and results may not accurately reflect reality, especially considering the significant impact of data quality as highlighted in [11]. In another work [3] authors used deep learning techniques to detect individuals with internet gaming-related disorders. They employed multichannel near-infrared spectroscopy to measure the brain activity of 40 individuals. Seven machine learning and deep learning models were then used to analyze and predict individuals with internet gaming disorders, trained on the multichannel near-infrared spectroscopy data. Results revealed that the convolutional neural network performed well in differentiating the two groups with a better prediction accuracy of 87.5%. However, criticisms were raised about the study, including the small sample size that may lead to biases and non-generalizable results. Additionally, using near-infrared spectroscopy alone

may not effectively measure individuals' brain activity with precision, necessitating the incorporation of other features to avoid limiting result scope. This concern aligns with findings that machine learning models often suffer from weak generalization when faced with small sample sizes and insufficient data diversity, particularly when external validation datasets are not used [12].

Another study by Aggarwal et al. [4] aimed to predict whether PlayerUnknown's Battlegrounds (PUBG) players were likely to develop video game disorders (IGD) and psychological disorders such as attention-deficit/hyperactivity disorder (ADHD) and generalized anxiety disorder (GAD). The authors used supervised machine learning models, including logistic regression and random forest, to analyze PUBG player data in Asian countries. The study also revealed a strong positive correlation between game statistics and IGD and ADHD, indicating the harmful effects of gaming. However, the study's sample size is limited to PUBG players in Asian countries, and the analysis is restricted to a specific genre of game. Therefore, the results may not be generalizable to other types of games and populations.

Finally, in the contribution proposed by [5] authors collected behavioral data using Google Forms and a standardized questionnaire related to gaming disorders. Feature engineering techniques were applied to find the most relevant characteristics to predict individuals most likely to develop a gaming addiction, and several supervised machine learning models were trained. However, it is worth noting that the study in question needs to incorporate additional data, such as individuals' medical data, to confirm these results.

3. Proposed Framework

3.1. Dataset Description

In the context of our article, we attempted to select an appropriate dataset for the purpose of predicting video game addiction. We found a dearth of datasets on gaming disorders. The datasets used in scientific articles [1,2,3,6] are private and inaccessible. Some of the accessible datasets are in text format [7] making them unsuitable for use with machine learning algorithms. Other datasets we came across [8] lacked the necessary target variable for applying supervised learning, which involves labeling individuals as addicts or non-addicts. Additionally, most research in the field of video game addiction focuses on using biomarkers to predict addiction disorders. However, our goal was to assess behavioral and demographic data to predict individuals at risk of developing video game addiction. To address this, we opted to use a dataset titled "Exploratory study of mental health among gamers in Gabon and Tunisia," which we obtained through Google Datasets [9]. Our choice was motivated by our desire to make an original contribution by predicting video game addiction solely based on behavioral and demographic data, without relying on individuals' medical data and biomarkers. Importantly, this dataset also includes the "Addicted gamers" column, which corresponds to the target variable. This variable comprises two values (0,1), representing the categories to which the gamers belong. This is essential for making predictions, making this dataset entirely suitable for the objectives of our study. The data sample used in our study consisted of video game players residing in two different countries, namely Gabon and Tunisia. The collected data included socio-demographic information such as age, gender, education, as well as behavioral variables related to gaming habits, mental health, and other relevant factors. The data exhibited variations depending on the measured variables. Variables related to gaming habits included elements such as the time spent on games, preferred game types, frequency of gaming sessions, and many other aspects. Variables related to mental health encompassed measures of video game addiction, anxiety, depression, and other characteristics.
Finally, Our study is the first to utilize this dataset for applying machine learning algorithms to analyze video game addiction, which adds originality to our work.

3.2. Features Engineering

The data sample used in our study consisted of video game players residing in two different countries, namely Gabon and Tunisia. The collected data included socio-demographic information such as age,

gender, education, as well as behavioral variables related to players' habits, mental health, and other relevant factors. The data showed variations based on the measured variables. Variables related to players' habits included elements such as time spent on games, types of preferred games, frequency of gaming sessions, and other factors. Mental health-related variables encompassed measures of video game addiction, anxiety, depression, and other characteristics. An essential step in this study involves verifying the equivalence of the data used. It is crucial to work with equivalent data, meaning data that has a similar number of instances in both classes of interest: individuals with video game disorders and healthy individuals. Unbalanced data can introduce bias into the results of machine learning models. To assess the data equivalence, we generated a tabular representation of instances for both classes. We observed a disparity in the number of instances between the class of individuals with video game disorders and the class of healthy individuals. This disparity can bias the results and lead to classification errors. To address this imbalance, we employed the ADASYN (Adaptive Synthetic Sampling) method. This method generates new artificial instances of the minority class by extrapolating the characteristics of existing instances. Its objective is to increase the number of samples in the underrepresented class and achieve a balance between the two classes. We chose the ADASYN method among several similar techniques because it is optimized to reduce the risk of overfitting. This step is motivated by the need to balance the number of instances in both classes, allowing the machine learning model to learn patterns from both classes to minimize biases.

4. Stacking Ensemble Learning Architecture

4.1. Architecture of the Proposed Stacking Ensemble Learning with three modules

The architecture we proposed is based on three complementary modules. The first module, Individual Model Training (IMT), supports the training of individual base models developed earlier, namely Random Forest, Decision Tree, XGBoost, and SVM. This module uses the 5-fold Cross-Validation method. Once all the different models are trained, they will be used to make predictions on the validation data.

The second module, Combined Predicted Values (CPV), is responsible for creating a new matrix with new features. The CPV module concatenates the predictions from the base models to form the matrix X of size (n×m), where 'n' represents the number of instances in the validation dataset for the meta-model, and 'm' represents the number of base models.

The third module, Meta-Model Training (MMT), handles the training of the meta-model. The Logistic Regression algorithm was chosen as the meta-model in our proposal because it allows for generating coefficients for each base model. To train the meta-model, we used the resulting X matrix from the CPV module and added the corresponding labels as targets. Then, the meta-model makes a final prediction on the test dataset.

The following Fig.1 illustrates the operating principle of the proposed architecture.

The architecture proposed in the previous section offers significant advantages in terms of prediction. Firstly, by combining the predictions of previously developed individual models, we can capture different aspects of the data, enhancing the robustness and performance of the meta-model. Each base model contributes its own expertise in detecting specific patterns or relationships in the data. However, this architecture has certain limitations, particularly regarding the interpretability of results for conducting a causal study. Indeed, video game addiction remains an important area of mental health, so it is essential for practitioners to understand the reasons that motivate the decisions made by the prediction models. By comprehending the underlying features and factors captured by the model, professionals in the field can make informed decisions about therapeutic practices to adopt and even customize these therapies to each individual addict's profile. In this context, and since we are combining predictions from multiple models, it can be challenging to fully understand and explain the decision-making process of Logistic Regression in the final prediction. This can be a challenge if the goal is to conduct an investigation rather than a predictive study.

Another limitation of stacking-based architectures becomes apparent as the data volume increases. When the dataset becomes very large, using the stacking method can lead to computational time issues.

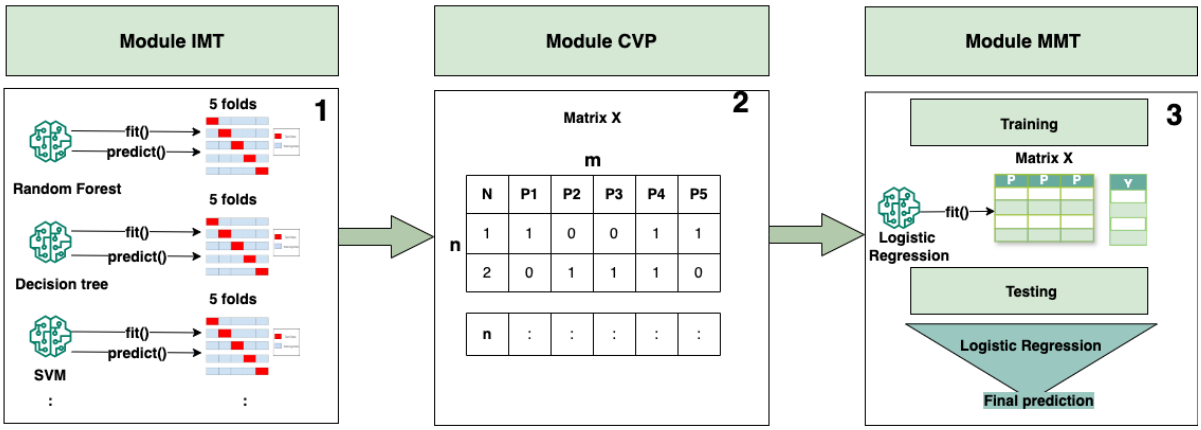

Figure 1: Architecture of the Proposed Stacking Ensemble Learning with three modules

The stacking architecture, therefore, requires training several base models on the training data, and then the models are used for predictions on the validation data. This significantly increases the complexity and computation time because each base model must be trained separately on large amounts of data. Furthermore, when dealing with large datasets, it can also be challenging to store and manage all the underlying model predictions for each instance in the validation dataset. This may require a substantial amount of memory capacity.

4.2. Architecture of the Proposed Stacking Ensemble Learning with Four Modules

To overcome the limitations of the architecture presented in section 4.1, we have added a fourth module that collaborates with the MMT module from the previous architecture, aiming to select the best model suited to the dataset from the base models. This new architecture offers significant advantages in scenarios with the previously mentioned limitations, which are execution time and interpretability. The added module, named Best Model Selection (BMS) (see Fig.2, facilitates the choice of the most performant model among the base models using the meta-model. This selection is made by using coefficients assigned by the Logistic Regression model to each base model based on their contributions to the final prediction. During the meta-model training process, coefficients are generated, which are values associated with the features in the X matrix (the predictions of the base models) in a meta-model (Logistic Regression). The initial values of these coefficients in the meta-model are typically set randomly during training. These coefficients determine the importance of each underlying model in the meta-model's final prediction. These coefficients are then adjusted to minimize the error between the meta-model predictions and the corresponding labels. Therefore, the meta-model seeks to find a balance between the contributions of the different base models. This process is performed automatically with the goal of optimizing the coefficients to obtain a meta-model capable of accurately predicting the target label based on the base model's predictions. In summary, the coefficients play a significant role in this process as they determine the relative weight of each underlying model in the final prediction. Base models that provide more information are assigned higher coefficients. Once the coefficients have been generated by the Logistic Regression model, we used the 'Argmax' function, which selects the index corresponding to the maximum value in a set of values to choose the base model with the highest coefficients. This means it has a significant impact on the final predictions of the meta-model. By adding the BMS module, we can reduce computation time by identifying the most performant model among the individual models. This allows us to focus resources on a single model rather than distributing them across multiple models, which can significantly accelerate the prediction process. Additionally, the selected model is considered a more straightforward alternative to interpret.

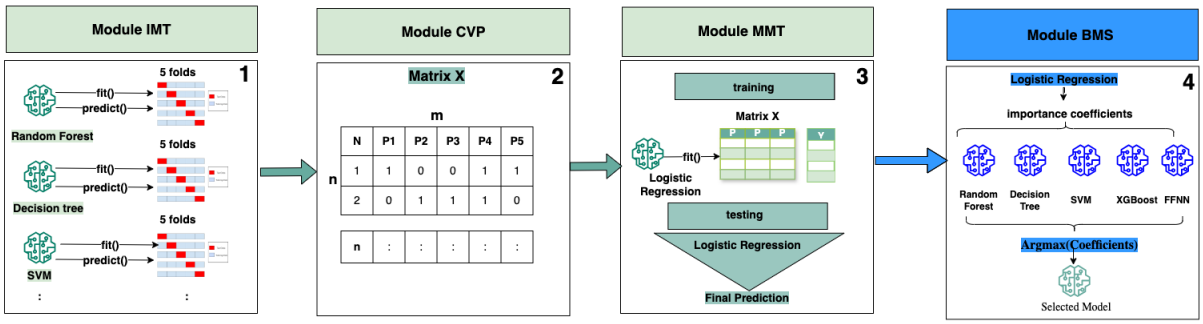

Figure 2: Architecture of the Proposed Stacking Ensemble Learning with Four modules

5. Evaluation Metrics

The evaluation metrics that were used in our study are as follows:
Recall: It is defined as the rate of True Positives (TP) or sensitivity, representing the proportion of correctly identified positive examples:

$$\text{Recall} = \frac{TP}{TP + FN}.$$

However, it is straightforward to achieve a high recall by predicting that all examples are positive . Therefore, this criterion cannot be used alone and is often associated with precision.

Precision: refers to the proportion of correct predictions among positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP}.$$

Accuracy: Also known as the rate of correct classification, it is an evaluation criterion used in classification that measures the proportion of examples correctly classified among all examples:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

6. Results and Discussion

6.1. Results of individual Machine Learning Models

The results of the machine learning models we developed individually are summarized in Table 1:

Table 1
The Results of Individual Machine Learning Models

Models	Accuracy	Recall	Precision
SVM	0.956	0.84	1
XGBoost	0.978	0.92	1
Decision Tree	0.978	0.92	1
Random Forest	1	1	1
Logistic Regression	0.978	1	0.926

By observing the performance metrics mentioned in this table, we found that SVM, XGBoost, and Decision Tree models seem to have relatively high scores with high precision and recall performances. However, it is important to note that in some cases, perfect scores may indicate the presence of overfitting issues.

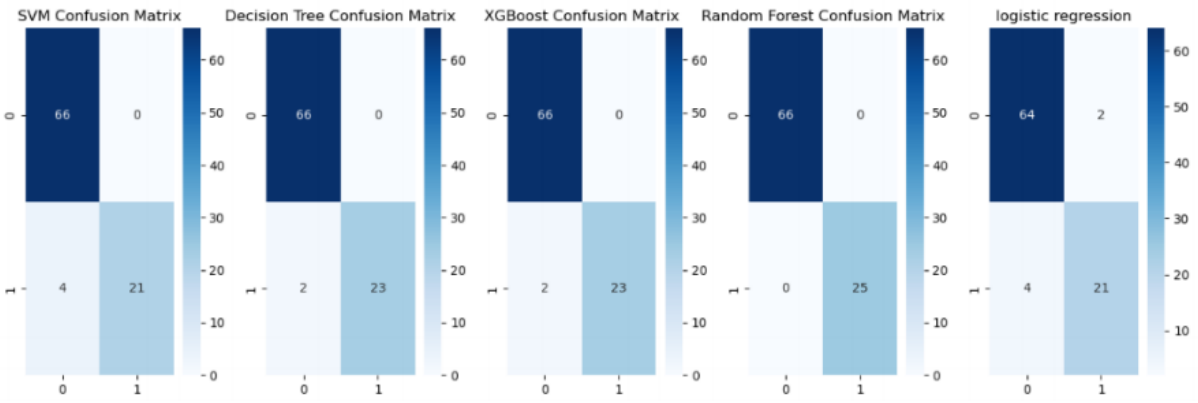

Figure 3: Confusion Matrix of Machine Learning Models

The Random Forest model achieved perfect scores for all evaluated metrics. This could be an indicator of overfitting. Indeed, when a model achieves precision, recall, and accuracy of 1.00, it may reveal that it has overly adapted to the specific training data, making it challenging to generalize well to new data.

The Logistic Regression model shows slightly lower precision at 0.926 compared to other models, suggesting some improvement in terms of generalization.

We have incorporated confusion matrices for the models, and the results are presented in the figure below. The scores from these confusion matrices show that models based on Decision Tree and XGBoost algorithms performed well by successfully classifying positive (True Positive, TP = 23) and negative (True Negative, TN = 66) instances. However, they produced two false positives (False Positive, FP = 2), indicating a failure to correctly classify two non-addict individuals as addicts. No false negatives (False Negative, FN = 0) were observed, and an increase in FN values is unfavorable in our study, as it means the model is not correctly identifying truly addictive individuals.

On the other hand, the model based on the Random Forest algorithm demonstrated excellent performance compared to the previous models. This model stands out with a high number of true positives (TP = 25) and the same value for true negatives (TN = 66). No false positives (FP = 0) or false negatives (FN = 0) were observed, demonstrating its high ability to correctly classify instances. However, it is important to note that this perfect performance could be an indicator of overfitting tendencies. Therefore, these results require a validation study to assess the model's generalization on new, unseen data.

Finally, the model based on the Logistic Regression algorithm obtained a lower number of true negatives (TN = 64) compared to SVM (TN = 66). This indicates that SVM is more effective in correctly predicting non-addictive individuals in the context of video game addiction. On the other hand, the results obtained with the deep learning model, represented in Fig.4 , show the curves of Recall, Precision, and Accuracy measures. The confusion matrix allows the analysis of the model's predictions based on actual categories. The scores of this matrix are depicted in Fig.5 : The corresponding confusion matrix reveals that the model correctly predicted 64 individuals not affected by video game addiction and 21 individuals affected by addiction. However, the confusion matrix also indicates that there were 2 cases of individuals with addiction incorrectly classified as non-affected and 4 cases of non-affected individuals wrongly classified as having addiction.

6.2. Results of Stacking Ensemble Architecture

We would like to emphasize that the performances of the individual base models within the stacking architecture have shown promising results in terms of accuracy, especially for XGBoost (0.944) and Random Forest (0.944). However, as highlighted in section 6.1, their tendency to overfit becomes apparent when data is limited. Figure 6 illustrates the performance of predictions from the basic models within the stacking architecture.

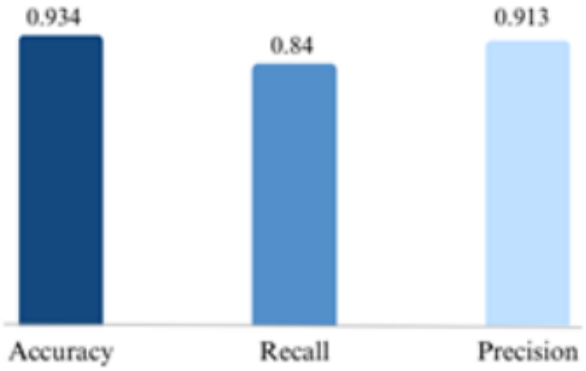

Figure 4: Results of FFNN

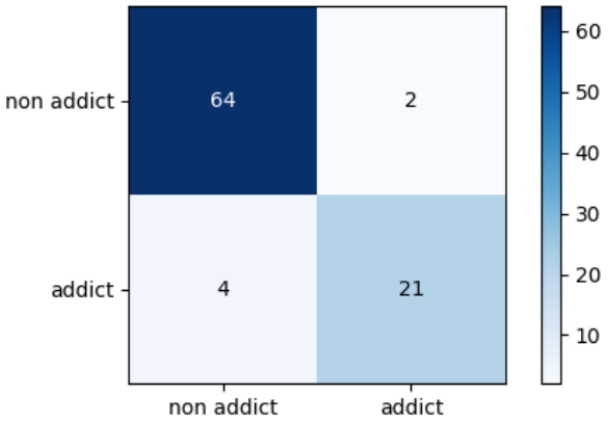

Figure 5: Confusion Matrix of FFNN

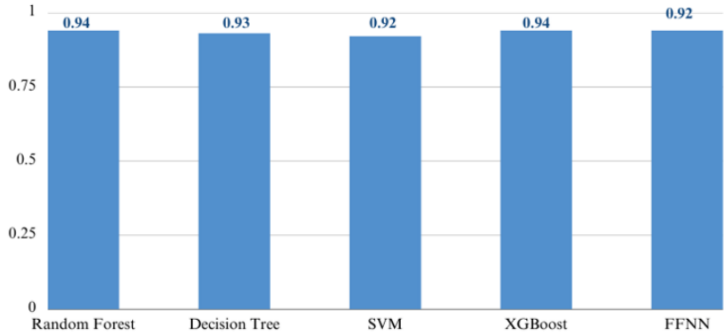

Figure 6: Results of Accuracy of base models inside Stacking Ensemble Model.

The meta-model achieved competitive performance compared to the base models (Accuracy: 0.95, Recall: 0.89, Precision: 0.87). Although the base models obtained highly accurate results, the meta-model combined their predictions in an optimal manner, thus producing a more realistic and robust prediction. For a better understanding of the performance of our meta-model (stacking-based architecture with three modules), Figure 7 presents the corresponding confusion matrix. Examining the values of this matrix, we observed that the model demonstrated a remarkable ability to predict non-addiction cases (True Negatives, TN = 237) as well as addiction cases (True Positives, TP = 51). This confirms the robustness of the stacking model in identifying players genuinely affected by video game addiction. On the other hand, the absence of null or high values for False Negatives (FN) and False Positives (FP) indicates that the model is not prone to bias. In other words, it does not tend to underestimate or overestimate the presence or absence of addiction in individuals. This characteristic is crucial for

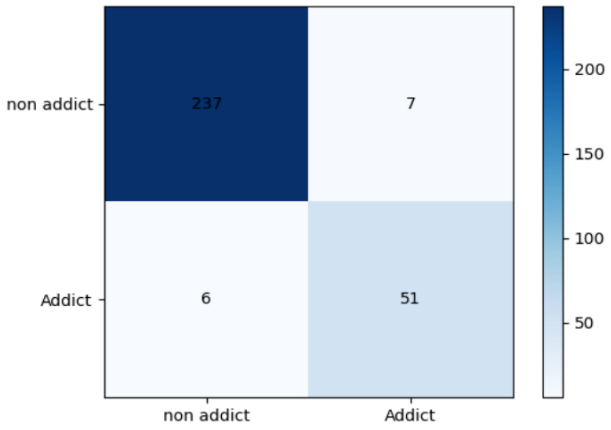

Figure 7: Confusion Matrix of Stacking Ensemble.

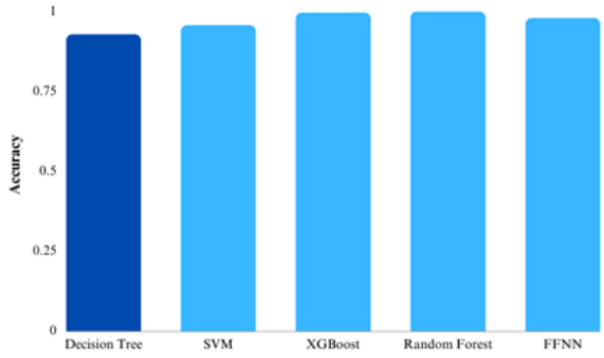

Figure 8: Best Model Selection.

ensuring realistic outcomes from this model.

Furthermore, during the process of selecting the most suitable model for the dataset among the base models, the meta-model chose the Decision Tree model, even though it was not the best in terms of Accuracy score. The selection was based on the importance coefficients generated by the Logistic Regression model. These coefficients highlighted the significant contribution of the Decision Tree model in the final prediction of the stacking model. Despite other models, such as XGBoost, having a high Accuracy score, the Decision Tree model was considered the most contributing according to the meta-model. These results emphasize the importance of not solely focusing on performance metrics. Figure 8 illustrates the selection of the best model among base models by the meta-model based on their coefficients.

7. Conclusion

In conclusion, this study employed a sophisticated approach, utilizing stacking ensemble learning, to predict gaming disorder. By comparing the results of individual models and the stacking ensemble, we aimed to identify the most suitable predictive model for our dataset. The comprehensive evaluation allowed us to discern the strengths and weaknesses of ensemble learning model and individuals models, in addition, the selection of an optimal model that strikes a balance between predictive accuracy and reduced complexity. Our findings underscore the importance of methodological considerations in predictive modeling for gaming disorder. The tailored selection of a model based on dataset characteristics and performance metrics is crucial for achieving accurate predictions while mitigating unnecessary complexity. This study contributes to the evolving field of predictive modeling for behavioral disorders and highlights the potential of stacking ensemble learning as an effective strategy for optimizing

predictive outcomes.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT for rephrasing and improving clarity of certain paragraphs, as well as Grammarly for grammar and spelling checks. All content generated or suggested by these tools was critically reviewed and edited by the authors. The author(s) affirm full responsibility for the accuracy, originality, and integrity of the final manuscript.

References

[1] X. Han, L. Wei, Y. Sun, Y. Hu, Y. Wang, W. Ding, and Y. Zhou, "MRI based radiomic machine-learning model may accurately distinguish between subjects with internet gaming disorder and healthy controls," *Brain Sciences*, vol. 12, no. 1, p. 44, 2021. doi:10.3390/brainsci12010044.

[2] H. Kim, L. Kim, and C. H. Im, "Machine-learning-based detection of craving for gaming using multimodal physiological signals: Validation of test-retest reliability for practical use," *Sensors*, vol. 19, no. 16, p. 3475, 2019. doi:10.3390/s19163475.

[3] Q. Wang, Y. Li, H. Su, N. Zhong, Q. Xu, and X. Li, "Deep neural network to differentiate internet gaming disorder from healthy controls during stop-signal task: a multichannel near-infrared spectroscopy study," *Biomedical Engineering/Biomedizinische Technik*, vol. 0, 2023. doi:10.1515/bmt-2023-0030.

[4] S. Aggarwal, S. Saluja, V. Gambhir, S. Gupta, and S. P. S. Satia, "Predicting likelihood of psychological disorders in PlayerUnknown's Battlegrounds (PUBG) players from Asian countries using supervised machine learning," *Addictive Behaviors*, vol. 101, p. 106132, 2020. doi:10.1016/j.addbeh.2019.106132.

[5] S. Islam, A. N. Tusher, M. S. Mia, and M. S. Rahman, "A Machine Learning Based Approach to Predict Online Gaming Addiction in the Context of Bangladesh," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-7, IEEE, October 2022. doi:10.1109/ICCCNT54827.2022.9984508.

[6] Data Station Social Sciences and Humanities, "Internet Gaming Disorder among Polish adolescents (Version 2.0) [Data set]," 2018. https://doi.org/10.17026/dans-x9a-xgem.

[7] Data Station Social Sciences and Humanities, "Internet Gaming Research Group 3: Internet Gaming Disorder (IGD) Psychopathology Culture [Data set]," 2020. https://doi.org/10.17026/dans-x3e-4452.

[8] Colombo University, "Data For: Internet Gaming Disorder: An Emerging Public Health Concern Among an Advanced Level Student Population from Colombo, Sri Lanka (Version 2) [Data set]," Mendeley, 2021. https://doi.org/10.17632/8r2jgm6ygh.2.

[9] Mendeley, "Exploratory study of mental health among gamers in Gabon and Tunisia (Version 4) [Data set]," 2020. https://doi.org/10.17632/c53rh2h435.2.

[10] A. Bouhadja and A. Bouramoul, "Beyond cocaine and heroin use: a stacking ensemble-based framework for predicting the likelihood of subsequent substance use disorder using demographics and personality traits," *International Journal of Computers and Applications*, vol. 45, no. 11, pp. 722–733, 2023. https://doi.org/10.1080/1206212X.2023.2273011.

[11] W. Abada and A. Bouramoul, "Exploring the Nexus of Mental Health Research and Data Synthesis: Role of Medical Data Quality in Machine Learning for Addiction Studies," in *2024 IEEE 7th International Conference on Advanced Technologies, Signal and Image Processing (ATSIP)*, vol. 1, pp. 465-470, IEEE, July 2024. doi:10.1109/ATSIP62566.2024.10638960.

[12] A. Bouhadja and A. Bouramoul, "A Review on Recent Machine Learning Applications for Addiction Disorders," in *2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, pp. 1-8, IEEE, October 2022.