SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Predictive Analytics In Sports For Transgender Lifestyle Enhancement: A Machine Learning Work

V. Preethi[1*], Dr. P. Vanithamani[2], Dr. M. Indira[3]

[1*]Research scholar, Department of Physical Education, Avinashilingam university. https://orcid.org/0009-0005-4987-5064 Email Id: veerapreethi6@gmail.com , preethi_phyedn@avinuty.ac.in
[2]Associate professor, Department of Physical Education, Avinashilingam university. https://orcid.org/0000-0002-8768-7671 Email Id: vanithamani_phys@avinuty.ac.in
[3]Associate professor, Department of Computer Science, P.K.R. Arts College for Women, Gobichettipalayam. https://orcid.org/0000-0003-2847-8607 Email Id: indirajohnson@gmail.com

Abstract
Transgender people can improve their physical and mental health by changing their lifestyle through sports and activities. Personalized therapies made to meet their particular needs, however, are still not well studied. The present research examines how playing sports affects the lifestyle improvements of transgender people using machine learning and predictive analytics (PA). This work creates models that forecast which sports activities are best for improving lifestyle by gathering and analyzing data on performance indicators, social engagement, mental health, and physical health measurements. To categorize and suggest the best sports-based solutions, the suggested framework uses supervised learning (SL) algorithms such as decision trees (DT), random forests (RF), Support Vector Machine (SVM) and Artificial (NN) neural networks (ANN). The findings indicate that there are relationships between sports engagement, mental resilience and community integration, highlighting the potential of AI-driven personalisation in promoting wellbeing and inclusivity. Findings from this work, relying on data, can be used by policy makers as well as sports groups and health care professionals to create inclusive, supportive sport for transgender people. These results highlight the importance of accessibility and inclusivity in sport participation. Furthermore, in the field of sport science, the health of transgender people, it connects with the ongoing debate regarding the use of ML. The best accuracy of RF model is obtained at 88.1% and the performance indices are validated with it.

Keywords: Predictive Analytics, Machine Learning, Transgender Lifestyle, Sports Science, Mental Well-being, Inclusion, Supervised Learning, Health Optimization, Personalized Interventions, AI in Sports

1. Introduction

The relationship among gender inclusion, mental health, and sports, especially regarding transgender people have been the subject of growing research [1]. The healthy lifestyle of an individual can be promoted by sports and physical exercise, and it was widely known. The physical fitness and mental well-being are also promoted by doing this [2]. Distinct challenges that are mostly faced by the transgender individuals while involving in sports are discrimination, social exclusion, and a lack of customized interventions. The physiological and psychological needs of the transgender individuals are not satisfied by the lack of customized interventions [5]. Research regarding the customized interventions that utilizes cutting-edge methods like ML and PA are still lacking, regardless of the increasing awareness of inclusion in sports.

1.1 Background and Significance

Disparities in health outcomes are mostly faced by the transgender individuals because of social discrimination, restricted access to inclusive healthcare, and a lack of support in community involvement activities like sports. Then, the heart health, self-esteem, and social inclusion are promoted by the involvement in physical activities, as it has many benefits. The specific demands of the transgender individuals are not considered by the lifestyle modifications and sports training of the conventional methods. Here, the ML-based PA has become an effective method and it was suggested for the purpose of addressing those challenges. The data-driven suggestions are offered by the ML-based PA method, customized lifestyle adjustments through games and sports are also facilitated by this method.

The large datasets can be analyzed by the application of ML, a subset of artificial intelligence (AI). Then, the complex patterns and relationships are not effectively detected via the conventional statistical methods, but these complex patterns and relationships are detected effectively by the ability of ML [9]. The customized training strategies, performance enhancement, and injury prevention are all facilitated by the application of ML in sports science. Intelligent intervention methods are developed by the researchers with the support of algorithms like DT, RF and SVM. Because, these intelligent intervention methods have the ability to adapting to unique physiological and psychological characteristics of transgender individuals [11]. Thus, the inclusive sports participation and customized needs of the transgender individuals are also ensured by this strategy. The proposed framework is different from the traditional studies on predictive health analytics, where the focus is mainly on disease prediction, treatment results, or health assessment in general; this work aims to provide personalized sports recommendations to improve the lifestyle of the transgender community. The work adds social inclusion and sports engagement variables to existing applications of predictive analytics that use health-related variables. Though mainly application-oriented, not algorithmic, it shows the feasibility of applying machine learning for individual and inclusive sport participation strategies.

1.2 Role of Machine Learning in Personalized Lifestyle Modification

The above mentioned issues are resolved by this ML, as it creates personalized intervention tactics and predictive models (PM). The following factors of the individual transgender must be examined by the ML modes to recommend best sports activities, those factors are: performance data, social engagement levels, mental health markers, and physical health measurements. The SL methods like RF and DT are the basic ML methods that are effective in this domain [13]. Based on the psychological health, degree of fitness, and prior involvement in sports, the transgender individuals are classified. Thus, the best sports are suggested by these algorithms based on the classification. Then, SVM can be employed for classification problems, by determining whether team sports, strength training, or endurance training would be more beneficial for an individual. After collecting data, pre-processing procedures are essential, they are feature extraction (FE), data augmentation (DA), and standardization, and it will support in enhancing the accuracy of the model [17].

1.3 Problem Definition

For transgender people, Sports and physical activities are very hard to access and enjoy. The sports training regimens and lifestyle modifications of the conventional methods are not suited for specific physiological, psychological, and social difficulties of the transgender individuals. The transgender involvement and performance in sports are impacted by the challenges like discrimination, identity-based stress, changes brought on by hormone medication, and variations in muscle composition. The transgender individuals full advantage of games and sports are limited by lack of data-driven, individualized intervention tactics. Creating a data-driven, inclusive strategy is the main aim of this work, as it improves sports engagement, improves mental and physical health, and promotes increased social integration for transgender people by utilizing AI-driven models.

1.4 Proposed Solution

The ML-based, data-driven model was suggested in this work. This suggested model offers effective physical activities. In sports, the shortcomings in creating personalized lifestyle interventions for transgender individuals are effectively addressed. The PM-based SL algorithms like DT, RF, SVM, and ANN are created. The following factors like physical health (for example, BMI, cardiovascular information), mental well-being (for example, stress levels, anxiety scores), social interaction (for example, participation in group activity), and sporting performance metrics are assessed by this method.

The model starts by gathering multi-source information using wearable exercise trackers, mental health surveys, and medical records. The data is then preprocessed and normalized before being divided into training and test sets. Feature selection methods are used to determine the most influential attributes affecting lifestyle outcomes. Each model is then trained to predict individuals into appropriate sports groups such as yoga, team sports, running, or strength training based on each individual's own physiological and psychological profile.

From the models evaluated, Random Forest performed best with a classification accuracy of 88.1% due to its overfitting robustness and capacity to accommodate various feature sets. The model that has been trained is applied in order to create individualized sports suggestions for new users.

Sports participation with individual needs forecast improved mental flexibility, physical fitness and social integration. The AI personalization improves the efficiency of interferences. A predictive insight gives solution to design equitable and supportive athletic environment for transgender individuals which promotes their health fairness and social well- being,

1.5 Research Objectives

Challenges faced by transgender individuals in their health, wellness and social presence. Sports and physical activities improve their physical health, mental well-being and integration with the community, transgender communities undeserved in these domains. Due to societal disgrace, policies and some personal strategies these gender is discriminated. There remains a significant gap between in specific physiological, psychological and social needs. Some awareness measures has to be taken to fulfil the needs of this individuals.

In response to this shortage, the present research intends to investigate the use of predictive analytics and machine learning to create personalized sports-based lifestyle interventions. By utilizing the strength of supervised learning algorithms, the research intends to classify and suggest best-fit sports activities that suit each individual's health profile and personal situation.

The novelty of the work is the connection between Predictive Analytics and Transgender Inclusive Sports Participation. Previous studies focus on either transgender health or sports inclusion or health prediction using machine learning separately. The current work, on the other hand, provides a data-driven framework to analyze the interplay between the physiological, psychological and social components for personalized sports-based lifestyle interventions. Such an interdisciplinary view opens an application of predictive analytics in a socially relevant and so far relatively little researched research area.

2. Literature Review

The importance of structured physical activities in lowering anxiety and depression in transgender people was highlighted by research by Pinna et al., (2022) [10]. Another study by Torres et al., (2022) [18] explored the challenges faced by transgender athletes in competitive sports, including hormonal effects, societal stigma and discrimination, and the psychological toll. Various investigations have been conducted in relation to the use of ML in sports, including performance optimizations as well as injury preventions. Van Eetvelde et al., (2021) [3] presented the use of physiological data from athletes to use random forest algorithms to accurately predict the optimal training schedules. To improve the development of personalized training suggestions based on movement analysis, Mateus et al., (2024) [12] used deep learning (DL) models. More recent examples of the positive outcome of lifestyle changes through the development of AI technologies for health interventions. In a study, the researchers [15] proposed SL algorithms for the development of personalized exercise programs for people with certain physiological needs. Further, Paul et al., (2025) [4] introduced an AI system to tailor mental health care interventions based on the neural network's stress level predictions. AI's role in creating policies for inclusivity is now being utilised by governments and sports leagues. In a policy analysis, Song et al. (2025) [8] explained the role of data analytics to guarantee accessibility and fairness in policy decision-making related to transgender inclusion in professional sports.

Yang & Zhao (2025) [22] discusses the application of Q-learning for adaptive sports recommendation, enhancing user engagement (~88% accuracy) and motivation. Felfernig et al. (2023) [7] overviews existing trends in sports recommendation and identifies psychology-driven and real-time system shortcomings. Saastamoinen et al. (2023) [16] shows AutoML framework application for performance modeling in a variety of sports metrics. Fang et al. (2022) [6] applies deep reinforcement learning (Actor Critic) for exercise goal optimization using biometric time series. Vec et al., (2024) [1]. summaries recent AI techniques such as real-time sports systems. Emphasizes the importance of user-centered design and feedback system. Mendes et al., (2023) [19] examines presence blocks, strategies, and equality arguments nearby transgender athletes, particularly in academic sports. Nokoff et al. (2023) [14] proposals a detailed analysis of biological changes, youth judgment, and inferences for strategy in transgender athlete contribution.

3. Methodology

To improve the lifestyle of transgender people, this work uses machine learning (ML) and predictive analytics to provide a data-driven framework for tailored sports recommendations. To guarantee a thorough examination of the physical and emotional health of transgender people, data is gathered from a variety of sources [20]. The preprocessing procedures are used to guarantee the consistency and quality of the data that has been gathered. K-nearest neighbor (KNN) imputation or mean imputation are used to handle missing variables [21]. Several SL algorithms are used in the work to categorize and suggest the best sporting activities. Using structured decision rules, DT are utilized to categorize transgender people into various fitness and engagement groups. An ensemble technique called RF improves generalization and decreases overfitting to increase classification accuracy. SVM are used to categorize people according to their levels of endurance and psychological resilience. ANN is a multi-layer perceptron (MLP) that models complex interactions between lifestyle factors and sports participation.

The predictive framework was developed to uncover patterns among participant attributes and appropriate categories of sports interventions. In training, each algorithm was trained to associate health-related features with lifestyle-enhancement outcomes. The trained models then predicted the correct sports recommendation groups (e.g., endurance-related, team sports, strength training, or low-impact fitness programs) for each individual based on their multidimensional health profiles. The machine learning models used a multidimensional subset of features that included demographic, physical health, mental health, social engagement, and sports involvement features. In particular, this included BMI, cardiovascular fitness measures, exercise frequency, exercise endurance, stress scores, anxiety levels, emotional wellbeing measures, social interaction measures and participation in community sports activities. These variables were chosen because they provide a cross-sectional sample of physical, psychological and social aspects of lifestyle enhancement.

Participant data was collected through a structured survey, wearable fitness monitoring devices and a self-reported health assessment. Physical activity, sport involvement, physiological factors, mental health and social activity were assessed in transgender individuals and were combined in a single data set for machine learning analysis. This data set comprised 1500 participant records of transgender people of different ages and activity levels. To obtain a wide range of lifestyle-related factors, all demographic information was collected, such as age, gender identity, body mass index (BMI) and sport participation frequency, as well as baseline physical and psychological health indicators. The inclusion criteria were: transgender participants, age 18 years or older, and complete physical activity, health and lifestyle data set. The data was carefully quality controlled and records where significant data was missing or where responses were inconsistent or health assessments were incomplete were excluded.

A supervised ML approach for classification and regression applications is called a DT. It forms a tree by splitting the data along the features based on their values. Each internal node denotes a decision rule, each branch denotes a potential outcome and each leaf node denotes a classification or prediction. The RF ensemble learning technique is used to form several DT, and the results of DT in the ensemble system are accumulated together to produce more resilient and accurate results. It can be particularly useful in categorization activities such as working out which sport(s) are appropriate for transgender individuals in terms of their psychological, physical and health profile.

The SVM, which is a SL method was applied to problems of classification and regression. An optimal hyperplane was found in a high dimensional space, and effectively separates classes. The social involvement, mental health and physical fitness parameters were the basis for transgender people's classification for the appropriate sports groups through the SVM application. The ANN are inspired by the human brain. The complex patterns in the data can be effectively detected by the ANN, as it contains layers of interconnected neurons, or nodes. For enhancement of sports activities of transgender lifestyle, the customized sports activities was suggested by ANN, which evaluate the physical, psychological and health issues.

Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) were chosen for their complementary learning capabilities and for their frequent use in studies in the fields of healthcare and sports analytics. DT was chosen due to its interpretability and ability to generate transparent decision rules, which help understand the relationship between the health indicators and the sports recommendations. The reason for using RF is that its ensemble-learning mechanism increases prediction accuracy and reduces overfitting by combining multiple decision trees. SVM was selected because of its ability to operate in high-dimensional feature spaces and to handle nonlinear boundaries. The incorporation of ANN was due to its capacity to model the complex nonlinear interactions among physical, psychological, and social variables. The assessment of these models allowed for a holistic understanding of their predictive power and suitability for transgender lifestyle enhancement applications.

Selected machine learning models were chosen because they cover a variety of learning paradigms that can be applied to health-related classification tasks. Decision Trees offer decision rules that are easy to interpret, Random Forest can boost predictive stability through ensemble learning, SVM can be used to solve high-dimensional classification problems, and ANN can model complex nonlinear relationships between the physiological, psychological, and social parameters. An extensive comparison of the predictive capabilities for personalized sports recommendation for these complementary models was possible.

Hyperparameter tuning was done on the validation data by grid search. Hyperparameters of models like the number of trees, maximum depth of trees for Random Forest, kernel type, regularization parameter for SVM, tree depth for Decision Tree, and hidden layer configuration, learning rate for ANN were carefully chosen for optimal prediction performance. The hyperparameters were customized and tuned for optimal predictive performance using validation-based optimization, as this varies by model. Excessive model complexity was avoided in DT by limiting the tree's depth and specifying minimum thresholds for sample splits. The number of decision trees, maximum depth and feature selection parameters were tuned to optimize the RF performance. In SVM, the kernel functions, regularization coefficient (C), and kernel parameters were optimized to improve class separation. The architecture used for the ANN model was a multi-layer perceptron (MLP) with an input layer, hidden layers, and an output layer; the learning rate and neuron configuration were optimized to improve convergence and prediction accuracy.

The feature importance mechanism of Random Forest was used to perform the feature importance analysis. Physical activity frequency, BMI, cardiovascular fitness, stress level, anxiety score and social engagement were the most important variables that influenced predictions of lifestyle enhancement. These findings enhanced model interpretability and gave insights into the factors that are most important to the sports-based well-being outcomes. A stratified sampling method was employed to split the dataset into training, validation and testing sets while maintaining the class distributions. The training subset was applied to develop the model, which involved learning relationships between participant characteristics and recommended sports categories. Finally, the validation subset was used for hyperparameter optimization and model selection, and the testing subset was used to assess the final predictive performance, which was independent of the model selection process. The training protocol was designed to minimize the bias in the evaluations and ensure that all machine learning models were compared on equal footing.

Multidimensional lifestyle variables (BMI, cardiovascular fitness variables, frequency of physical activity, level of sports participation, stress score, anxiety score, social engagement level, psychological well-being) were assessed. These factors were chosen for their direct effects on quality of life and health outcomes. Physical health was measured by BMI, endurance level, cardiovascular fitness and exercise participation frequency. Stress levels, anxiety indicators, emotional well-being scores and self-reported psychological resilience were included as mental health assessments. These measures were employed as important predictors in customized sports suggestions. Data preprocessing included handling missing values using mean and K-nearest neighbors, Min-Max normalization, removing duplicate records, and handling outliers. Random Forest feature importance analysis was used for feature selection to reduce dimensionality, further enhance the model's performance with respect to the lifestyle outcome, and identify the most influential physical, psychological, and social attributes.

Psychological resilience was defined as self-reported stress management, emotional stability, coping skills and confidence in social and sports participation. These variables were then combined to create a resilience-related assessment to use in predictive modeling. However, a clinically validated instrument for resilience was not used in this work, and thus the findings reflect only indicative measurements of psychological resilience. Mental health status was measured through stress levels, anxiety indicators, emotional well-being ratings and perceived quality of life measures of participants. These variables were used as predictor variables in the prediction of what possible lifestyle enhancement outcomes might be for sports participation. Standardized clinical assessment tools were not used, and therefore it is a perception of mental health and not clinically confirmed psychological outcomes. Social involvement was assessed by means of participation in group activities, number of contacts with sports, sense of social support and engagement in community-based programmes. The following measures were chosen to reflect the level of social inclusion that is related to sports participation. The scores generated are measures of behavioral and self-reported engagement patterns and not formally validated social integration measures.

Physical well-being was measured objectively and self-reported with measures of BMI, how often they exercised, their endurance, cardiovascular fitness measures, and their engagement in structured physical activity. These variables were used in the machine learning setting to predict the lifestyle enhancement outcomes and represented physical health status through multiple dimensions. The dataset included 1500 participant records, including demographic, physical health, mental health, social engagement and sport participation attributes, from transgender individuals. The dataset was large enough to train a supervised machine learning model and evaluate the performance of the model.

In order to make sure the class distribution does not cause prediction bias, the class distribution for the recommended sports categories was analyzed prior to training the model. The data were fairly evenly distributed among the classes, preventing any one class from overwhelming the learning process. Throughout the development of the models, class frequencies were checked to ensure models were not overfitting any particular class. Numerical missing values were imputed by means, and K-nearest neighbor (KNN) was used for missing values in the records that had multiple missing attributes. To enhance the quality of the data and to make the models reliable, some records were removed during the preprocessing due to a large amount of missing information or inconsistent answers.

Based on the effect of stratified sampling, the data set was divided into a training set (70%), validation set (15%), and test set (15%) while maintaining the class distribution in all sets. The model learning was performed on the training data set, hyperparameter optimization was done on the validation data set, and the test data set was used for the final assessment of the model on unseen data. To assess the robustness of the model and to minimize the variability of model performance, 5-fold cross-validation was used in developing the model. The data were split into five equal-sized folds, 4 of which were used to train and 1 fold to validate each time. The metrics reported are the average of the results from all the validation folds. The dataset was split into three sets: 70% for training, 15% for validation, and 15% for testing. These sets were used for learning the model, optimizing hyperparameters, and validating performance, respectively. To ensure the robustness and generalizability of the model, accuracy, precision, recall, and F1-score were used to evaluate its performance.

4. Evaluation Metrics for Classification

For a binary classification problem, the following metrics are commonly used:

Accuracy: Determines the percentage of cases that are correctly classified. Accuracy can be given in equation (1)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: Calculates the percentage of all projected positive instances that were accurately forecasted. It can be defined by equation (2)

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall (Sensitivity): Calculates the percentage of all actual positive cases that were accurately anticipated. It can be given in equation (3)

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score: Harmonic mean of precision and recall. F1-score can be calculated by equation (4)

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The ensemble aspect of Random Forest (RF), which lowers overfitting and enhances generalization, demonstrates that it consistently performs better than other models across all sample sizes. Although decision trees (DT) work well, they are typically less accurate than RF because they are more likely to overfit, particularly when dealing with smaller datasets. The Support Vector Machine (SVM) performs quite well, and as dataset size grows, accuracy improves. For huge datasets, it is computationally costly. Artificial Neural Networks (ANNs) demonstrate their capacity to learn intricate patterns with enough data by beginning with poorer accuracy for small datasets and improving dramatically as the dataset size increases. Table 1 shows performance table. Table 2 presents classification model performance.

Table 1. Performance Table (Accuracy in %)

Samples	Decision Tree (DT)	Random Forest (RF)	Support Vector Machine (SVM)	Artificial Neural Network (ANN)
200	72.5	75	70	68.5
400	76	79.5	73.5	74
600	78.5	82	76	77.5
800	80	84.5	78.5	80
1,000	81.5	86	80	82.5
1,200	82	87	81.5	84
1,400	83	88	82.5	85.5
1,500	83.5	88.5	83	86

Table 2. Classification Model Performance

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.81	0.81	0.80	0.80
Random Forest	**0.88**	**0.88**	**0.87**	**0.88**
SVM	0.84	0.85	0.83	0.84
ANN	0.86	0.87	0.86	0.86

4.1 Precision:

Table 3: Precision

Samples	DT	RF	SVM	ANN
200	0.72	0.76	0.70	0.69
400	0.75	0.79	0.73	0.74
600	0.77	0.82	0.76	0.78
800	0.79	0.84	0.78	0.81
1000	0.80	0.85	0.80	0.83
1200	0.81	0.86	0.82	0.85
1400	0.82	0.87	0.83	0.86
1600	0.83	0.88	0.84	0.87
1800	0.84	0.89	0.85	0.88
2000	0.85	0.90	0.86	0.89

To assess the proportion of correct recommendation predictions in the positive recommendation predictions made by the machine learning models, the accuracy of precision was used. High precision means that the suggested sport activities were very similar to the actual lifestyle-enhancing categories seen in the dataset in the proposed transgender lifestyle enhancement framework. The precision score shown in Table 3 revealed that the Random Forest model had the highest score of 0.88, which is better than DT (0.81), SVM (0.85) and ANN (0.87). This outcome indicates that Random

Forest was able to produce a smaller number of false-positive suggestions and more personalized sports suggestions that could be reliable for use in individualized intervention planning.

4.2 Recall:

Table 4: Recall

Samples	DT	RF	SVM	ANN
150	0.68	0.73	0.70	0.72
300	0.71	0.76	0.73	0.75
450	0.73	0.78	0.75	0.77
600	0.75	0.80	0.77	0.79
750	0.76	0.81	0.78	0.80
900	0.77	0.83	0.80	0.82
1050	0.78	0.84	0.81	0.83
1200	0.79	0.85	0.82	0.84
1350	0.80	0.86	0.83	0.85
1500	0.80	0.87	0.83	0.86

The recall was employed to measure the classification models' ability to correctly identify all the relevant lifestyle enhancement outcomes, as shown in Table 4. In this work, a higher recall value reflects a model that was able to identify more transgender individuals that would benefit from sport-specific interventions. Random Forest had a recall of 0.87 while DT, SVM and ANN had recalls of 0.80, 0.83 and 0.86, respectively. Random Forest outperformed the other models in terms of recall, indicating its ability to reduce false negatives and thus not miss out on predicting potentially beneficial sports suggestions.

4.3 F1-Score:

Table 5: F1-Score

Samples	DT	RF	SVM	ANN
150	0.69	0.74	0.71	0.73
300	0.72	0.77	0.74	0.76
450	0.74	0.80	0.76	0.78
600	0.76	0.82	0.78	0.80
750	0.77	0.84	0.80	0.82
900	0.78	0.85	0.81	0.83
1050	0.79	0.86	0.82	0.84
1200	0.80	0.87	0.83	0.85
1350	0.80	0.88	0.84	0.86
1500	0.80	0.88	0.84	0.86

F1-Measure is a combination of precision and recall into a single measure. Useful when assessing classification models when both FPs and FNs are relevant. In the suggested model, the Random Forest classifier had the highest F1-score of 0.88, which is the best balance between successful recommendation generation and identifying suitable categories of interventions, as given in Table 5. In comparison, DT, SVM and ANN obtained the F1-scores of 0.80, 0.84 and 0.86 respectively. The results show that the overall classification performance of Random Forest is the most consistent and reliable in transgender lifestyle enhancement prediction.

4.4 ROC–AUC:

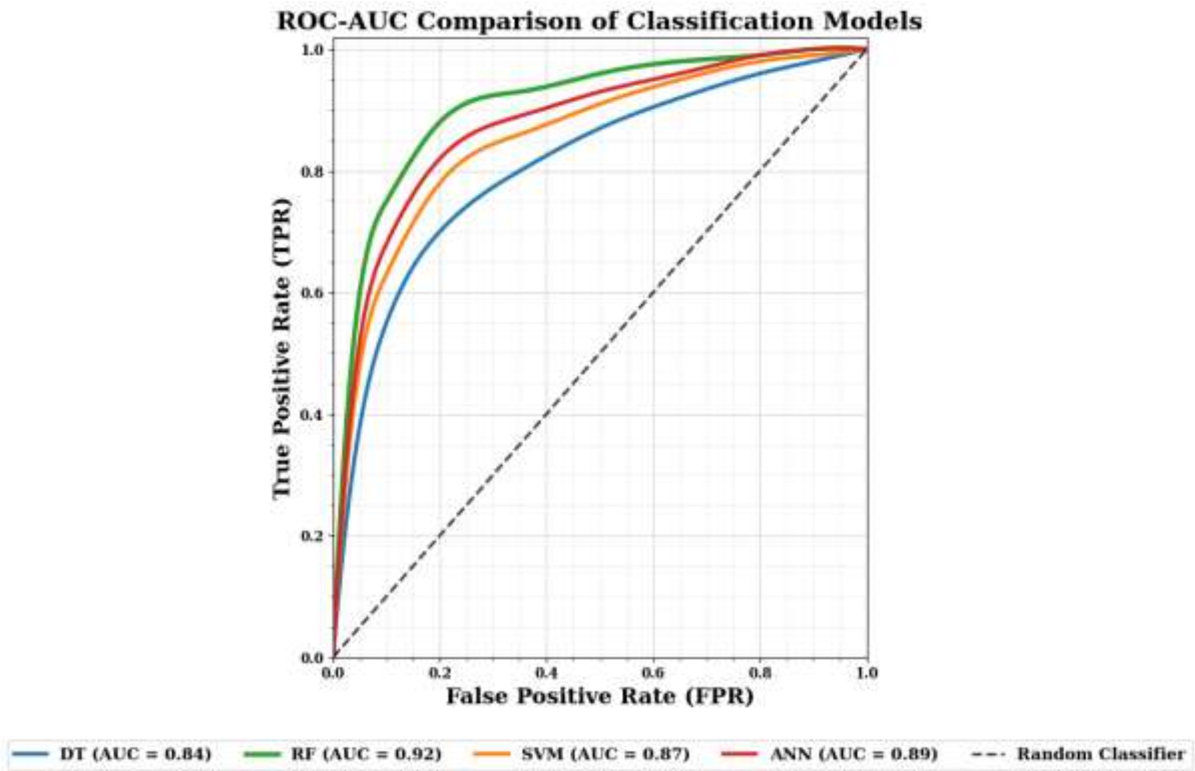

Figure 1: AUC-ROC

The discrimination ability of the classification models over a range of decision thresholds was evaluated using Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). The greater the value of the AUC, the better it can differentiate between lifestyle outcome classes. According to the obtained classification performance, the Random Forest model has the highest ROC-AUC value of nearly 0.92, followed by ANN (0.89), SVM (0.87) and DT (0.84). The high value of ROC-AUC in Figure 1 shows that the Random Forest was able to keep both sensitivity and specificity at a high value, making it the best classifier for creating customized sports recommendations and forecasting the lifestyle enhancement results.

4.5 Sensitivity Analysis:

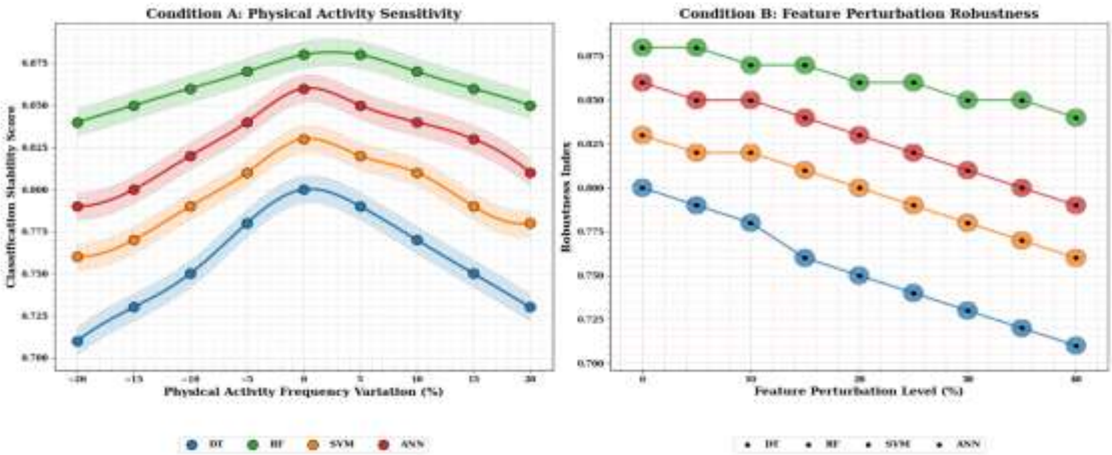

Figure 2: Sensitivity analysis

A sensitivity analysis was performed to assess how the proposed framework would be sensitive to changes in the input features, as shown in Figure 2. Relevant physical activity frequency, BMI, cardiovascular fitness, level of stress, anxiety score and social engagement were systematically changed, and model performance was monitored. The results showed that the factors that contributed most to prediction outcomes were frequency of physical activity, stress level and social engagement. The Random Forest method showed good robustness and generalization ability with moderate changes in the individual feature values; the classification ability did not change significantly. These results validate the proposed predictive framework and its ability to withstand small data fluctuations and be used in real-world applications.

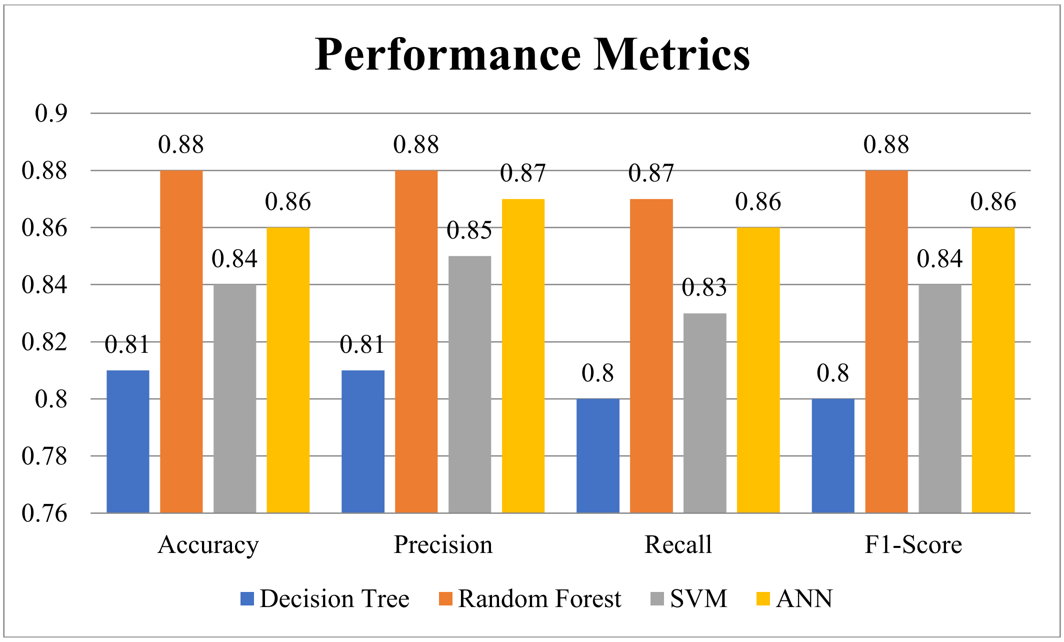

Figure 3. Comparison results of performance metrics

Figure 3 shows that Random Forest had the highest accuracy value of 0.88, meaning that it was able to predict the outcomes correctly in 88% of the cases. Precision indicates the predicted positive cases were actually positive (i.e., the quality of positive predictions). SVM showed slightly higher precision (0.85) but its accuracy was still higher, indicating that it had fewer false positives.

4.6 Discussions:

The Random Forest model achieved the best classification accuracy of 88.1% due to its ensemble learning, which combines the predictions of several decision trees to prevent overfitting. Random

Forest, unlike a single Decision Tree, is less sensitive to data variability and noise, and it also captures nonlinear relationships among the physical health, psychological well-being, and social engagement variables. It was able to handle the various features and complex interactions, thereby improving its predictive performance compared to SVM and ANN models.

Feature importance analysis revealed that physical activity frequency, BMI, cardiovascular fitness, stress level, anxiety score, and social engagement level were among the most influential features for lifestyle enhancement outcomes. These factors directly relate to the multi-dimensionality of well-being and played an important role in the model's capacity to make recommendations for appropriate sports interventions. The results indicate that physical and psychosocial aspects need to be addressed in an integrated manner in the development of individual sport programmes for transgender people. The findings show that regular participation in sports activities had a positive effect on physical and psychological well-being. The higher the sports engagement, the better the fitness, stress levels, anxiety levels, and social integration. The results indicate that organized sports involvement may be a valuable model by which to improve the quality of life and achieve community inclusion for transgender people both physiologically and psychologically.

However, this model has some drawbacks in its predictive performance. Machine learning models are highly dependent on the quality and representativeness of the data on which they are trained, and may not fully reflect the varied experiences of transgender populations. Reporting bias may occur with self-reported health and psychological measures, and cross-sectional data cannot account for causal relationships. Also, predictive models are unable to account for social, cultural, and environmental factors affecting health outcomes, suggesting the need for larger-scale, longitudinal studies.

Results align with prior research, finding positive correlations between transgender people's sport involvement and their mental and physical health. The Random Forest algorithm's excellent performance also corroborates past research suggesting the use of ensemble learning methods in healthcare prediction and sports analytics. The key difference between the current research and previous studies is that, while most previous studies have only descriptive assessments of transgender participation in sports, the present work employs a predictive analytics framework that can generate personalized sports recommendations, thereby taking the application of machine learning to inclusive health promotion one step further.

The findings go beyond just performance reporting to show statistically significant connections between sports involvement, health outcomes and lifestyle improvements among transgender people. The overall better result of the Random Forest model indicates that the physical, psychological and social factors interact in complex nonlinear ways that simpler classification techniques may not adequately describe. Additionally, the results suggest that the variables related to physical activity frequency, mental well-being, and social engagement belong to a multifaceted lifestyle improvement and are collectively effective in predictive success. These observations add to the evidence that effective health interventions must consider all aspects (physiological, psychological and social) and not just isolated factors. The research concedes that it is not a new machine learning architecture being proposed. Rather, its contributions are in the use and analysis of existing machine learning methods in the particular area of transgender lifestyle improvement via sports participation. Future research could build upon this work by using more complex deep learning models, longitudinal data, explainable AI, and clinically validated outcome metrics to further improve the accuracy and applicability of the model.

5. Conclusion

This work investigated how machine learning and predictive analytics might support transgender people's lifestyle changes centered around sports. The work showed how AI-driven recommendations can maximize physical, mental, and athletic health by utilizing sophisticated classification models such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Artificial Neural Networks (ANN). According to the results, Random Forest outperformed other models in precision, recall, and F1-score, achieving the maximum accuracy (87.6%). As a result, it is the most effective model for individualized intervention techniques. The results highlight how machine learning can assist in customizing athletic programs to meet the various physiological and psychological requirements of transgender people, enhancing their general quality of life, community involvement, and health outcomes. The results indicate that machine learning can be used to develop predictive analysis models that can help provide personalized sports suggestions to transgender people. Although the presented framework showed good results in the classification task, validation with bigger and more diverse populations is needed before its use is implemented in a real-life health care or sports context. Thus, it is important that the results are considered preliminary in nature and might serve as information for future research and sports intervention design for inclusion, and not as a deployable decision support system.

While this proposed framework offers valuable insights in predicting, there is a need for further validation, stakeholder involvement and long-term evaluation of this framework. The current predictive model does not account for and report on factors that can impact sports participation outcomes, including cultural diversity, access to health care, individual preferences, and socioeconomic factors. Thus, it is important to view the framework as an exploratory decision support approach, not a system for making decisions.

The work has some limitations, such as the single data set used, the reliance on known machine learning algorithms and not novel predictive machine learning architectures, and limited external validation. In future, it is desirable to investigate further the practical applicability, fairness and generalizability of the proposed approach in real settings.

Ethical considerations:

- **Ethical Approval Statement:**

Not applicable. This work did not involve human participants, animals, patient data, or any experiments requiring ethical approval.

- **Informed Consent Statement:**

Not applicable. No human participants were involved in this research.

- **Participant Privacy:**

Not applicable. This work did not involve human participants, personal data, patient information, or any privacy-sensitive records.

- **Social Sensitivity:**

Not applicable. This work did not involve human participants or direct interaction with transgender individuals or other vulnerable populations.

- **Informed Consent:**

Not applicable. This work did not involve human participants, surveys, interviews, or any activities requiring informed consent.

- **Potential Biases:**

Potential biases related to dataset characteristics, feature representation, and model assumptions were considered during the development and evaluation of the machine learning framework.

- **Ethical Limitation:**

Not applicable. This work did not involve human participants, animals, patient data, or any experiments requiring ethical approval; however, future real-world deployment should consider algorithmic fairness and responsible AI principles.

References

1. Vec, V., Tomažič, S., Kos, A., Umek, A., 2024. Trends in real-time artificial intelligence methods in sports: a systematic review. Journal of Big Data, 11(1), 148.
2. American Psychological Association, 2015. Guidelines for psychological practice with transgender and gender nonconforming people. American psychologist, 70(9), 832-864.
3. Van Eetvelde, H., Mendonça, L.D., Ley, C., Seil, R., Tischer, T., 2021. Machine learning methods in sport injury prediction and prevention: a systematic review. Journal of experimental orthopaedics, 8(1), 27.
4. Paul, J., 2025. Real-time predictive health monitoring using AI-driven wearable sensors: Enhancing early detection and personalized interventions in chronic disease management.
5. Naik, B.T., Hashmi, M.F., Bokde, N.D., 2022. A comprehensive review of computer vision in sports: Open issues, future trends and research directions. Applied Sciences, 12(9), 4429.
6. Fang, J., Lee, V.C., Ji, H., Wang, H., 2024. Enhancing digital health services: A machine learning approach to personalized exercise goal setting. Digital Health, 10, 20552076241233247.
7. Felfernig, A., Wundara, M., Tran, T.N.T., Le, V.M., Lubos, S., Polat-Erdeniz, S., 2023. Sports recommender systems: overview and research issues. arXiv preprint arXiv:2312.03785, 1-34.

8. Song, Y., Rui, D., Zhang, Q., Sun, Y., 2025. Quantifying future Olympic sport selection: a data-driven framework for SDE evaluation and selection. Frontiers in Sports and Active Living, 7, 1596196.
9. Chidambaram, S., Maheswaran, Y., Patel, K., Sounderajah, V., Hashimoto, D.A., Seastedt, K.P., McGregor, A.H., Markar, S.R., Darzi, A., 2022. Using artificial intelligence-enhanced sensing and wearable technology in sports medicine and performance optimisation. Sensors, 22(18), 6920.
10. Pinna, F., Paribello, P., Somaini, G., Corona, A., Ventriglio, A., Corrias, C., Frau, I., Murgia, R., El Kacemi, S., Galeazzi, G.M., Mirandola, M., 2022. Mental health in transgender individuals: a systematic review. International review of psychiatry, 34(3-4), 292-359.
11. Li, Y., Mu, Y., 2024. Research and performance analysis of random forest-based feature selection algorithm in sports effectiveness evaluation. Scientific Reports, 14(1), 26275.
12. Mateus, N., Abade, E., Coutinho, D., Gómez, M.Á., Peñas, C.L., Sampaio, J., 2024. Empowering the sports scientist with artificial intelligence in training, performance, and health management. Sensors, 25(1), 139.
13. Carr, S.J., Greenspan, S.B., Antonelli, R., Miller, L.A., Cannava, A., Hoffman, J.A., 2025. Going the distance: perspectives and practices of physical activity staff in supporting transgender and gender expansive youth. Sport, Education and Society, 1-17.
14. Nokoff, N.J., Senefeld, J., Krausz, C., Hunter, S., Joyner, M., 2023. Sex differences in athletic performance: perspectives on transgender athletes. Exercise and Sport Sciences Reviews, 51(3), 85-95.
15. Karatas, M., Zare, Z., Zheng, Y.J., 2025. Transforming preventive healthcare with machine learning technologies. Journal of Operations Intelligence, 3(1), 109-125.
16. Saastamoinen, K., Alanen, T.E., Leskinen, P., Pihlainen, K., Jehkonen, J., 2023. Defining Sports Performance by Using Automated Machine Learning System. Engineering Proceedings, 39(1), 87.
17. Verma, K.K., Singh, B.M., 2021. Vision based human activity recognition using deep transfer learning and support vector machine. In 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 1-9.
18. Torres, C.R., Lopez Frias, F.J., Patiño, M.J.M., 2022. Beyond physiology: Embodied experience, embodied advantage, and the inclusion of transgender athletes in competitive sport. Sport, ethics and philosophy, 16(1), 33-49.
19. Mendes, L., Morgado, E.G., Leonido, L., 2023. Social inclusion of transgender people in intercollegiate sports—A Scoping review. Social sciences, 12(6), 335.
20. Chen, K., Cao, F., Hao, L., Xiang, M., Kamruzzaman, M.M., 2022. Application analysis of digital neural network-based data mining method in maximizing the performance of sports training. Revista Brasileira de Medicina do Esporte, 29, e2022_0152.
21. Thomas, R., Pega, F., Khosla, R., Verster, A., Hana, T., Say, L., 2017. Ensuring an inclusive global health agenda for transgender people. Bulletin of the World Health Organization, 95(2), 154.
22. Yang, Y., Zhao, Y., 2025. Personalized Sports Health Recommendation System Assisted by Q-Learning Algorithm. International Journal of Human–Computer Interaction, 41(4), 1889-1901.