Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

SVM Optimization for Autism Spectrum Disorder Classification: A Comparison of PCA, PSO, and Grid Search

Aura Choirun Nisa[1]*, Basuki Rahmat[2], Achmad Junaidi[3]

[1,2,3] Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur
21081010173@student.upnjatim.ac.id[1], basukirahmat.if@upnjatim.ac.id[2], achmadjunaidi.if@upnjatim.ac.id[3]*

Abstract

Autism Spectrum Disorder (ASD) is a developmental condition impacting communication and socialization, often manifesting in distinct behaviors. Early detection and timely intervention are crucial for improving the quality of life for individuals with ASD. This research aims to develop an ASD risk classification model using the Support Vector Machine (SVM) algorithm across three age groups: children, adolescents, and adults. To optimize model performance, Principal Component Analysis (PCA) was used for dimensionality reduction, while Particle Swarm Optimization (PSO) and Grid Search were employed for parameter tuning. The study sought to identify the most effective combination of these techniques for autism prediction. Evaluation results indicated that SVM with Grid Search optimization, without PCA, yielded the best performance, achieving 98.2% accuracy and an AUC of 0.997 at an 80:20 data split. Furthermore, Grid Search demonstrated greater computational efficiency compared to PSO. The findings suggest that the integration of SVM and Grid Search offers a promising, accurate, and efficient approach for the early detection of autism.

Keywords: *SVM; PCA; PSO; GridSearch; Autism Spectrum Disorder (ASD)*

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that causes difficulties in communication, behavior, language, and social interaction. Individuals with autism often exhibit distinct behaviors, such as repetitive tendencies in specific activities, and present a highly diverse spectrum of symptoms [1]. Research on autism, such as that conducted by Apeksha [2] and Patel [3], focuses on identifying language patterns in individuals with autism to improve early diagnosis and understanding of its symptoms. Autism symptoms in children typically emerge between the ages of 2 and 3 years, such as difficulties in communication and social interaction [4]. In adolescents and adults, symptoms are often more challenging to recognize as they have developed coping strategies. However, they struggle to build social relationships like neurotypical individuals and also display repetitive behaviors, making them appear different from individuals without autism [5].

Bawa's research [6]details efforts to build a classification system using machine learning techniques to differentiate children, adolescents, and adults with Autism. The research results indicate that the application of machine learning methods, with the Support Vector Machine (SVM) being the best model optimized with Grid Search for hyper-parameters, can improve model accuracy up to 99.55%. Nevertheless, the uneven distribution of labels can still introduce bias into the model.

Based on the research in the preceding paragraphs, SVM is a classifier model that can be effectively used to classify autism. SVM is a supervised learning algorithm, frequently employed in pattern recognition and data analysis, with the ability to separate data into distinct categories with high accuracy [7]. SVM is generally used to classify both linear and non-linear data by utilizing kernel functions to transform low-dimensional data into higher dimensions, enabling linear separation between two classes using a hyperplane. Kernel functions allow SVM to handle more complex data, such as in real-world applications, with hyperplanes varying according to the number of data dimensions [8]. Furthermore, SVM has proven effective in the early detection and diagnosis of various health conditions, including issues related to drug abuse, liver disorders, respiratory problems, hepatitis, and autism spectrum disorder [9].

One technique that can be used to optimize SVM is Principal Component Analysis (PCA) for reducing data dimensionality. Additionally, parameter optimization techniques such as Grid Search are employed. This technique has the advantage of simplicity for hyperparameter optimization by evaluating every possible combination and selecting the best one. For example, in research classifying Arabic sentiment data, Grid Search successfully optimized model parameters and improved performance in handling various classification cases [10]. Moreover, recent research on handling high-dimensional data in stunting cases in Samarinda City also shows that Grid Search is effective for parameter optimization in SVM models with various kernels. This method can find the best parameter combination, improve SVM model performance, and produce more accurate predictions [11].

Other parameter optimization techniques, such as Particle Swarm Optimization (PSO), can also be considered. PSO is an optimization algorithm introduced by Kennedy and Eberhart in 1995 [12], inspired by the social behavior of animal groups, where particles move within the search space to find the optimal solution by updating their position and velocity through iterations until the desired solution or iteration limit is reached [13][14]. Both Grid Search and PSO can enhance model accuracy. Previous research indicates that the use of PSO can be more efficient in exploring large parameter spaces, making it a faster and more effective alternative in the search for optimal solutions [15][16].

This research aims to compare the performance of SVM in classifying ASD using three optimization techniques: PCA, PSO, and Grid Search. The data used includes autism screening data for various age groups. The performance comparison will be based on metrics such as accuracy, precision, recall, F1-Score, and AUC, with the goal of finding the most effective optimization method for improving SVM model performance.

2. Research Method

The research methods used in this study are data collection, data preprocessing, data dimension reduction to two types of ratios, namely 80:20 and 70:30, followed by model implementation and model testing, and ended by model evaluation to analysis to draw final conclusions. The model is built in the Gcolab environment. The following on Fig. 1 is an overview of the research flow

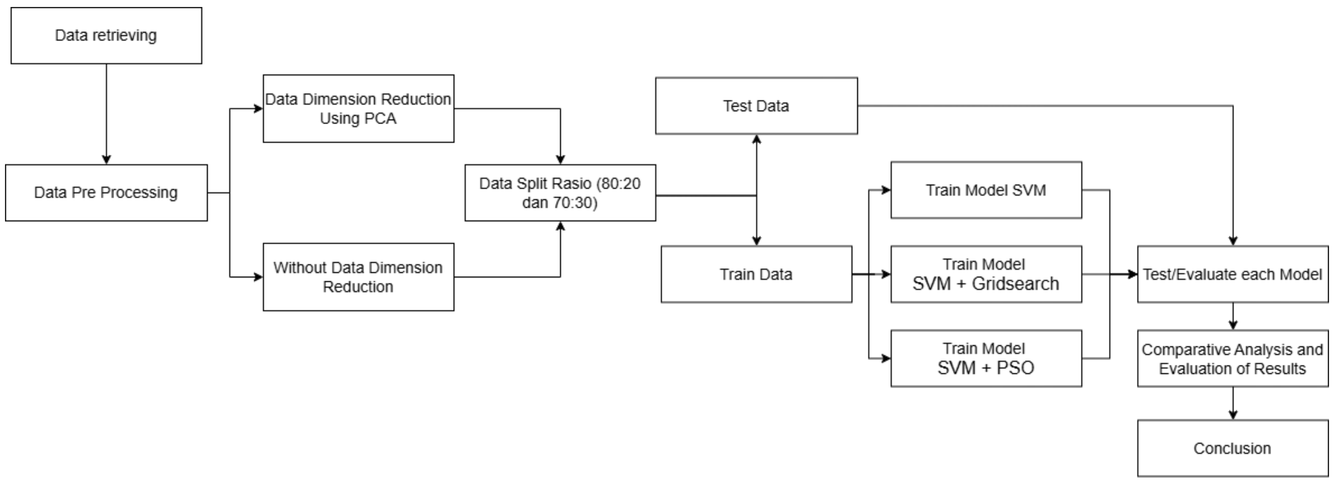

Fig. 1: Research Workflow

2.1. Datasets

The data used in this study was obtained from the autism screening dataset, which is available at the UCI Machine Learning Repository. This dataset consists of three files, each based on a specific age group (children, adolescents, and adults), and each file is in CSV format. The details of the datasets utilised are delineated in Table 1.

Table 1: Datasets Overview

Characteristic	Children	Adolescents	Adults
Number of Data	292	104	704
Number of Atributes	20	20	20
Missing Value	4	0	2
Age Ranges	4-11 Years	12-17 Years	18-65 Years

The total amount of data utilised in the study was approximately 1100 entries. The dataset itself includes attributes as outlined in table 2, such as :

Table 2: Datasets Atributes

ID Atribut	Atribut	Tipe
1	Age	Number
2	Gender	String
3	Ethnicity	String
4	Jaundice	Boolean (yes or no)
5	Family Members with PDD	Boolean (yes or no)
6	Who Completed the Test	String
7	Country of Residence	String
8	Have Used Screening Application Before	Boolean (yes or no)
9	Type of screening test	Integer (0,1,2,3)
10 -19	Answers to 10 Screening Questions	Binary (0, 1)
20	Screening Score	Integer

The screening question is the AQ-10, otherwise known as the Autism Spectrum Quotient -10. The AQ-10 is a screening tool that was developed by researchers in Cambridge in 2001 for the purpose of detecting early indications of autism in children, adolescents, or adults. The 10 questions evaluate aspects of an individual's social behaviour associated with autism [17].

2.2. Data Preprocessing and Data Dimension Reduction

The preprocessing stage in this study consists of handling missing values using standard deviation, encoding non-numeric data to numeric, and rectifying the labels of age ranges or categories that are not yet neat. This is to ensure that the data can be processed by the classification model at a later stage. The data was then reduced using PCA with a 95% threshold, resulting in 17 attributes from the initial 20 attributes. This process guarantees that the projected data form retains the majority of the information from the original dataset in a lower dimension [18]. The two forms of data were then separated into two ratios, 70:30 and 80:20 for training and test data, resulting in four types of data that could be utilised.

2.3. Model Implementation and Scenario Testing

Three model scenarios are employed, firstly, standard SVM without parameter optimisation, secondly, SVM with parameter optimisation using Gridsearch, and thirdly, SVM with parameter optimisation using PSO. Following the completion of the model, a total of 12 test scenarios are conducted. This scenario is obtained from training and testing each type of model on the four types of data utilised. Subsequently, an analysis will be conducted of the evaluation results of the scenario, which will then be compared with the effect of each optimisation. This will be undertaken in order to ascertain the most optimal combination from several perspectives, including accuracy and computation time.

3. Result and Discussion

Tests have been conducted using a total of three scenario models: namely, standard Support Vector Machine, Support Vector Machine with Gridsearch, and Support Vector Machine with PSO.

Tests conducted on the Multi-Class Classification, where labels are separated into child_yes and child_no categories, as well as adolescent_yes and adolescent_no categories, and adult_yes and adult_no categories, allow for a more in-depth and thorough investigation of whether or not a person has ASD. The overall analysis of each model was conducted based on the following metrics: accuracy (Acc), recall (Rec), precision (Prec) and f1-score (F1 Score). The relevant equations are given below:

$$Acc = \frac{N_C}{N_P} \tag{1}$$

where N_C and N_P are the number of correct predictions and the total number of predictions or total data, respectively.

$$\text{Rec}=\frac{T_p}{(T_p+F_N)} \tag{2}$$

$$\text{Prec}=\frac{T_p}{(T_p+F_p)} \tag{3}$$

$$\text{F1-Score}=\frac{2*\text{Prec}*\text{Rec}}{\text{Prec}+\text{Rec}} \tag{4}$$

where T_p, F_p and F_N represent the number of positive data that are correctly predicted positive, the number of negative data that are incorrectly predicted positive, and positive data that are incorrectly predicted negative, respectively. These figures are derived from the prediction outcomes of the modelling technique employed.

Furthermore, additional metrics such as the Area Under the Curve (AUC) are utilised to evaluate the model's capacity to differentiate between classes. Area under the curve (AUC) is obtained from the ROC (Receiver Operating Characteristic) curve, which is formed from the true positive rate (recall) against the false positive rate at various probability threshold values. The probability values in question are derived from the model's prediction output, rather than the final label (0 or 1), and consequently, the ROC and AUC assess the model's performance in all possible decisions. The closer the AUC value is to 1, the greater the model's capacity to differentiate between positive and negative classes.

3.1. Classification Model Evaluation Results

The following in table 3 is the evaluation data of model testing results for the use of data divided by a ratio of 70:30 training and test data.

Table 3: Evaluation Results for Model using 70:30 Rasio Data

Model Combination	PCA threshold%	Data Rasio	Accuracy	Precision	Recall	F1 Score	AUC
SVM	Non PCA	70:30	0,979	0,981	0,979	0,979	0,985
SVM+GridSearch	Non PCA	70:30	0,979	0,981	0,979	0,979	0,993
SVM+PSO	Non PCA	70:30	0,979	0,981	0,979	0,979	0,995
SVM	95	70:30	0,904	0,907	0,904	0,903	0,965
SVM+GridSearch	95	70:30	0,904	0,909	0,904	0,904	0,965
SVM+PSO	95	70:30	0,907	0,908	0,907	0,906	0,967

The results of the evaluation matrix demonstrate that the accuracy of models utilising unreduced data or PCA produces equivalent evaluation metrics. This is due to the fact that the parameters utilised in the training of the three models are identical. It is notable that the parameters obtained by the evaluation results are consistent. In the case of models that utilise reduced data via PCA, a substantial decline is observed in comparison to the preceding three models. Nevertheless, the implementation of SVM + PSO with PCA data yields superior outcomes in terms of evaluation metrics when compared to alternative model types. This finding is consistent with the research conducted by Raghunath, which demonstrates that the integration of PCA and PSO in SVM can enhance the classification accuracy of intrusion detection systems when compared to conventional methods [19]. In a similar vein, a study by Atteia revealed that the PCA-PSO-SVM approach yielded higher diagnostic accuracy for blood cancer compared to the use of PCA or PSO methods in isolation [20].

Furthermore, when evaluated from the AUC value of each model, all models utilising non-PCA data exhibited elevated AUC, with SVM+PSO demonstrating the highest performance, followed by SVM+GridSearch and standard SVM. Conversely, when the data is reduced using PCA, the AUC exhibits a marginal decline; nevertheless, the SVM+PSO model continues to demonstrate superior performance, attaining an elevated AUC value in comparison to the other two models. This finding indicates that the combination of PCA and PSO can potentially maintain overall classification performance, particularly in terms of the trade-off between data dimensionality and model predictive ability.

The subsequent section of Table 4 presents the evaluation data pertaining to the model test results, wherein the utilisation of data has been divided into a ratio of 80:20, encompassing training and test data, respectively.

Table 4: Evaluation Results for Model using 80:20 Rasio Data

Model Combination	PCA threshold%	Data Rasio	Accuracy	Precision	Recall	F1 Score	AUC
SVM	Non PCA	80:20	0,982	0,983	0,982	0,982	0,996
SVM+GridSearch	Non PCA	80:20	0,982	0,983	0,982	0,982	0,997
SVM+PSO	Non PCA	80:20	0,982	0,983	0,982	0,982	0,997
SVM	95	80:20	0,900	0,905	0,900	0,900	0,965
SVM+GridSearch	95	80:20	0,878	0,887	0,878	0,880	0,960
SVM+PSO	95	80:20	0,901	0,874	0,900	0,875	0,928

Based on the test results in the above table, the results are broadly similar to those in the model using 70:30 data. For the model using 80:20 data without dimensionality reduction via PCA, accuracy increases, whereas for the model using PCA, accuracy decreases compared to the previous 70:30 ratio. This demonstrates that the impact of PCA on model performance can vary depending on the data sharing ratio. This finding is consistent with research by Shaufee, which showed decreased accuracy at an 80:20 ratio compared to a 70:30 ratio. Even when different parameters were used, the accuracy decreased, indicating that the data sharing ratio affects model performance, particularly when PCA is used [21].

Additionally, examining the AUC value of each model reveals that all models using non-PCA data exhibit high AUC values. SVM+PSO and SVM+GridSearch exhibit similar values, followed by standard SVM. These higher AUC values are consistent with the results of the study by Saputra et al., which showed that the SVM-PSO model performed better than ordinary SVM, particularly in terms of AUC, after parameter optimisation [22]. Meanwhile, in data reduced by PCA, the AUC decreased significantly in all SVM + PSO models. A decrease in the AUC of the SVM+PSO model of about 39% between the 70:30 and 80:20 ratios indicates that a larger data ratio can reduce predictive ability. However, SVM and SVM+GridSearch did not experience a decrease in the AUC. This suggests a trade-off between dimensionality reduction and model predictability: SVM+PSO performance degrades.

3.2. Performance Comparison of Each Type of Model

For further comparison, the accuracy results obtained previously are compared with those of the model using PCA-reduced data with different thresholds (90 and 99) to gain more insight into the effect of the PCA dimension reduction threshold on the model's performance. The following line diagram is based on the split ratio and type of data used. The diagrams are grouped based on the model used.

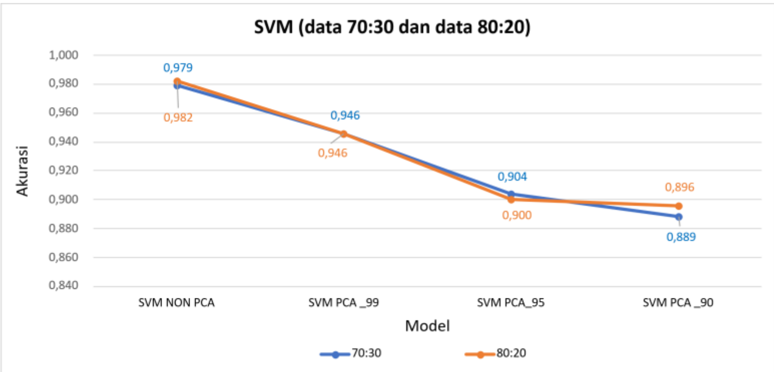

Fig. 2: Accuracy performance line diagram of the SVM model

Figure 2 shows similar results for the two lines. The model's accuracy is shown using standard data without dimensionality reduction. At certain points, the 70:30 data ratio is more accurate than the 80:20 ratio, particularly when the data is reduced using PCA with a threshold of 90. The model without PCA achieves the highest accuracy, at 0.979 for the 70:30 ratio and 0.982 for the 80:20 ratio. Using PCA with a threshold of 99% reduced the accuracy to 0.946 for both ratios. Further reductions in the threshold to 95% and 90% resulted in a further decrease in performance, reaching 0.904 and 0.889 for the 70:30 ratio and 0.900 and 0.896 for the 80:20 ratio, respectively. This demonstrates that the lower the PCA threshold, the more information is lost, resulting in decreased accuracy.

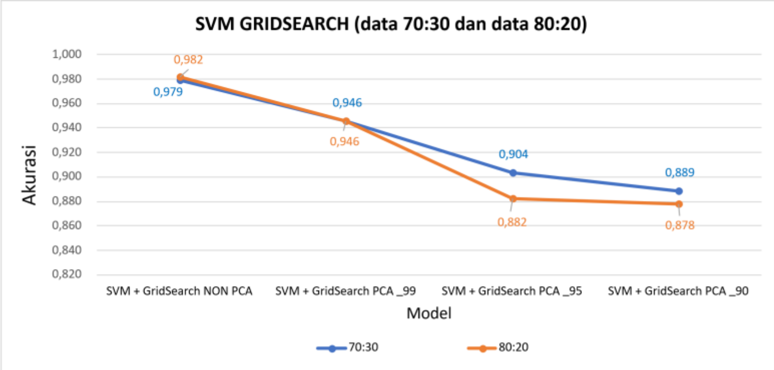

Fig. 3: curacy performance line diagram of the SVM Gridsearch model

A similar pattern of performance degradation is also shown in the accuracy results of the SVM model with Gridsearch in Figure 3. The highest accuracy is still found in the model without PCA (0.979 and 0.982), whereas the accuracy with the 99% PCA threshold decreases to 0.946. The 95% and 90% thresholds resulted in a more significant drop: even with an 80:20 data ratio, the accuracy only reached 0.882 and 0.878 respectively. This reinforces the conclusion that Gridsearch works optimally with non-PCA data, while using PCA considerably impacts accuracy despite more efficient computing time. Research by Alaika and Alamsyah also supports these findings: the combination of SVM and correlation-based feature selection improves accuracy [23]. However, based on the results obtained, the use of dimension reduction techniques such as PCA needs to be carefully reconsidered.

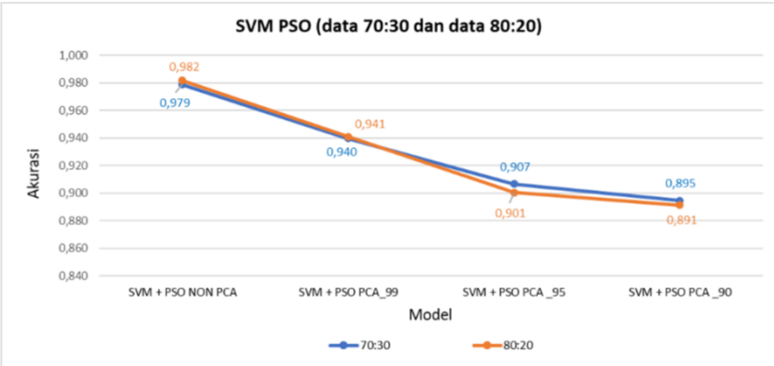

Fig. 4: curacy performance line diagram of the SVM PSO model

Figure 4 also shows the previous pattern, where the highest accuracy is achieved by not using PCA (0.979 and 0.982). Using a 99% PCA threshold resulted in a slight decrease (0.940 and 0.941), whereas using a 95% threshold resulted in a more pronounced decrease (0.907 and 0.901), as did using a 90% threshold (0.895 and 0.891). However, compared to previous GridSearch models, accuracy increased by about 6–9% at the 95% threshold and 14–18% at the 90% threshold for the 70:30 ratio, and by about 6–9% at the 95% threshold and 14–18% at the 90% threshold for the 80:20 ratio. This demonstrates that the ratio significantly impacts the application of PSO in determining SVM classification parameters.

Judging from the accuracy results obtained by using either PCA with different thresholds or no PCA in the previous description, the best order of models for classification is Standard SVM, followed by SVM + PSO and finally SVM + GridSearch. Standard SVM demonstrates stable and accurate performance in various conditions, whereas SVM + PSO improves accuracy at several thresholds. Meanwhile, SVM + GridSearch did not significantly increase accuracy, especially at the 95% and 90% thresholds, thus occupying last place. Despite the variation in accuracy among the models, other evaluation metrics must be considered to make a final decision.

3.3. Comparison of Model Computational Time

In addition to the previously analysed performance metrics, other metrics such as computation time are also important considerations in determining the best model combination. For this reason, the line chart below is presented to compare the computational time of each tested model, whether using PCA or not.

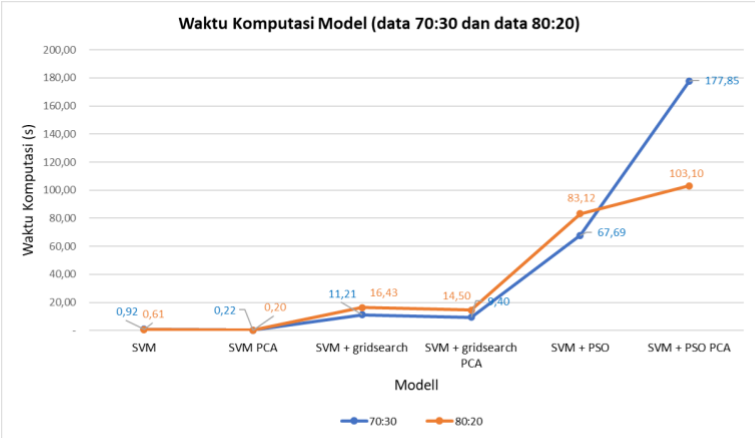

Fig. 5: Model computation timeline diagram

As can be seen from the graphs in Figure 5 above, the SVM model without PCA achieves the fastest computation times: 0.92 seconds for the 70:30 ratio and 0.61 seconds for the 80:20 ratio. Using PCA with SVM reduces the computation time further to 0.22 seconds for the 70:30 ratio and 0.20 seconds for the 80:20 ratio, demonstrating the efficiency of dimensionality reduction. However, combining SVM with optimisation techniques such as GridSearch and PSO significantly increases the computation time. The SVM + GridSearch model without PCA takes 11.21 seconds for the 70:30 ratio and 16.43 seconds for the 80:20 ratio, whereas the computation time is slightly reduced to 9.40 seconds and 14.50 seconds, respectively, when PCA is used. The computation time in the SVM + PSO model shows an even more drastic increase. Without PCA, it reached 67.69 seconds (70:30) and 92.92 seconds (80:20). With PCA, it increased to 177.85 seconds (70:30) and 103.10 seconds (80:20). This demonstrates that combining PSO and PCA can result in substantial computational overhead.

Overall, the standard SVM model offers the best balance of accuracy and time efficiency. While the SVM + PSO model improved accuracy in certain PCA scenarios compared to GridSearch, it came with significantly higher computation time. Conversely, SVM + GridSearch tends to produce lower accuracy than PSO at low thresholds and is not as computationally efficient as standard SVM or PCA. PCA has been proven to accelerate the computational process by reducing dimensionality, which aligns with Jumato's findings that high dimensional complexity affects the efficiency and accuracy of the model used [24]. Additionally, Manik's study found that, while the combination of SVM and PSO can improve accuracy, it requires more computational time than other optimisation methods [25].

4. Conclusion

Based on the research conducted on Support Vector Machine (SVM) optimisation for autism spectrum disorder (ASD) classification using principal component analysis (PCA), particle swarm optimisation (PSO) and grid search, several conclusions were obtained. Firstly, the application of PCA with a 95% threshold was proven to significantly reduce the number of features and speed up the computational process, albeit with a slight decrease in model performance compared to using the original data without reduction. GridSearch was found to be more efficient than PSO in terms of both computation time and accuracy stability. The best configuration was obtained for SVM without PCA, optimised using GridSearch, achieving an accuracy of 98.2% and an AUC of 0.997 at a data ratio of 80:20. PSO produced almost equivalent accuracy but required much longer computation time, rendering it less efficient. Overall, Gridsearch is recommended as the most effective parameter optimisation method for ASD classification using SVM, particularly when time efficiency is a primary concern.

Acknowledgement

I gratefully acknowledge the contributions of all those who supported this research. Special thanks are due to the faculty members who provided guidance and the mentors for their invaluable assistance and insights. I also deeply appreciate the cooperation of participating organizations and individuals. Finally, I thank everyone who offered their time and expertise to this work. I hope that this report proves beneficial, and I welcome constructive feedback for future development..

References

[1] F. G. Winarno, *Autisme dan Peran Pangan*. Jakarta: PT Gramedia Pustaka Utama, 2013.
[2] K. Apeksha, S. Hanasoge, P. Jain, and S. S. Babu, "Speech Perception in Quiet and in the Presence of Noise in Children with Autism Spectrum Disorder: A Behavioral Study," *Indian Journal of Otolaryngology and Head & Neck Surgery*, vol. 75, no. 3, pp. 1707–1711, 2023, doi: 10.1007/s12070-023-03721-5.
[3] S. P. Patel *et al.*, "A profile of prosodic speech differences in individuals with autism spectrum disorder and first-degree relatives," *J Commun Disord*, vol. 102, Mar. 2023, doi: 10.1016/j.jcomdis.2023.106313.
[4] N. Aisyah Amalia and S. Tjiptorini, "ANALISIS ANAK DENGAN GANGGUAN SPEKTRUM AUTISME (GSA) DI SEKOLAH LUAR BIASA ULAKA PENCA BERDASARKAN DSM V," 2024.
[5] Z. Muna, Mahdalevia, U. Fatiya, and Hijriati, "Analisis Karakteristik Anak Bekrebutuhan Khusus (Autisme) di Sekolah PAUD Harsya Ceria Banda Aceh," *Journal Of Early Childhood And Islamic Education*, vol. 2, no. 2, Jun. 2024.
[6] P. Bawa, V. Kadyan, A. Mantri, and H. Vardhan, "Investigating multiclass autism spectrum disorder classification using machine learning techniques," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Jun. 2024, doi: 10.1016/j.prime.2024.100602.
[7] F. Lumbanraja, F. Lufiana, Y. Heningtyas, and K. Muludi, "IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PEDERITA DIABETES MELLITUS," *Jurnal Komputasi*, vol. 10, Apr. 2022, doi: 10.23960/komputasi.v10i1.2940.

[8] H. Syah and A. Witanti, “ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM),” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 5, no. 1, Feb. 2022, doi: 10.47080/simika.v5i1.1411.

[9] R. Dasmasela, B. P. Tomasouw, and Z. A. Leleury, “PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MENDETEKSI PENYALAHGUNAAN NARKOBA,” *Parameter: Jurnal Matematika, Statistika dan Terapannya*, vol. 1, no. 2, Oct. 2022.

[10] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, “Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis,” *Informatics*, vol. 8, no. 4, Dec. 2021, doi: 10.3390/informatics8040079.

[11] Siti Muawwanah, Taghfirul Azhima Yoga Siswa, and Wawan Joko Pranoto, “Model Optimasi SVM-GSBE dalam Menangani High Dimensional Data Stunting Kota Samarinda,” *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 3, pp. 1246–1258, Jul. 2024, doi: 10.32493/jtsi.v7i3.41545.

[12] B. Adiguzel Mercangoz, *Applying Particle Swarm Optimization New Solutions and Cases for Optimized Portfolios: New Solutions and Cases for Optimized Portfolios*, 1st ed. Springer Cham, 2021. doi: 10.1007/978-3-030-70281-6.

[13] R. K. Samala, “Particle Swarm Optimization,” in *Swarm Intelligence - Recent Advances and Current Applications*, M. A. Aceves-Fernández, Ed., Rijeka: IntechOpen, 2023. doi: 10.5772/intechopen.107156.

[14] T. Wiantara and S. Samsuryadi, “Attribute selection and weighting using C4.5 and Particle Swarm Optimization for dementia disease classification ,” Thesis (Undergraduate), Sriwijaya University., 2021.

[15] X. Chen, S. Xu, S. Li, H. He, Y. Han, and X. Qu, “Identification of architectural elements based on SVM with PCA: A case study of sandy braided river reservoir in the Lamadian Oilfield, Songliao Basin, NE China,” *J Pet Sci Eng*, vol. 198, Mar. 2021, doi: 10.1016/j.petrol.2020.108247.

[16] S. Samantaray, A. Sahoo, and A. Agnihotri, “Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River Basin,” *MethodsX*, vol. 10, Jan. 2023, doi: 10.1016/j.mex.2023.102060.

[17] L. J. Francis, A. B. Sailer, V. J. Payne, and U. McKenna, “Engaging the AQ10 to Predict Professional Burnout or Poor Work-Related Psychological Wellbeing Among Anglican Clergy in Wales,” *J Relig Health*, vol. 63, no. 2, pp. 1661–1676, Apr. 2024, doi: 10.1007/s10943-024-02006-7.

[18] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D’Enza, A. Markos, and E. Tuzhilina, “Principal component analysis,” *Nature Reviews Methods Primers*, vol. 2, no. 1, p. 100, 2022, doi: 10.1038/s43586-022-00184-w.

[19] M. P. Raghunath *et al.*, “PCA and PSO based optimized support vector machine for efficient intrusion detection in internet of things,” *Measurement: Sensors*, vol. 37, Feb. 2025, doi: 10.1016/j.measen.2024.101806.

[20] G. Atteia, R. Alnashwan, and M. Hassan, “Hybrid Feature-Learning-Based PSO-PCA Feature Engineering Approach for Blood Cancer Classification,” *Diagnostics*, vol. 13, no. 16, Aug. 2023, doi: 10.3390/diagnostics13162672.

[21] L. H. Shaufee, H. Jantan, U. Fatihah, and M. Bahrin, “Polycystic Ovary Syndrome (PCOS) Prediction System Using PSO-SVM,” *Journal of Computing Research and Innovation*, vol. 9, no. 1, 2024, doi: 10.24191/jcrinn.v9i1.

[22] D. Saputra, W. S. Dharmawan, and W. Irmayani, “Performance Comparison of the SVM and SVM-PSO Algorithms for Heart Disease Prediction,” *International Journal of Advances in Data and Information Systems*, vol. 3, no. 2, pp. 74–86, Nov. 2022, doi: 10.25008/ijadis.v3i2.1243.

[23] L. Alaika, “Optimization of Accuracy to Autism Spectrum Disorder Identification for Children Using Support Vector Machine and Correlation-based Feature Selection,” *Journal of Advances in Information Systems and Technology*, vol. 4, no. 1, 2022, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/jaist

[24] Jumanto *et al.*, “Optimizing Support Vector Machine Performance for Parkinson’s Disease Diagnosis Using GridSearchCV and PCA-Based Feature Extraction,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 38–50, Feb. 2024, doi: 10.20473/jisebi.10.1.38-50.

[25] A. Manik, E. B. Nababan, and Tulus, “IMPROVED SUPPORT VECTOR MACHINE PERFORMANCE USING PARTICLE SWARM OPTIMIZATION IN CREDIT RISK CLASSIFICATION,” *JUTIF : Jurnal Teknik Informatika*, vol. 3, no. Vol. 3 No. 6 (2022): JUTIF Volume 3, Number 6, December 2022, pp. 1739–1746, Dec. 2022, doi: https://doi.org/10.20884/1.jutif.2022.3.6.615.