VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

INTELLIGENT SYSTEM FOR EARLY DETECTION OF DIABETES MELLITUS IN CHILDREN USING SUPPORT VECTOR MACHINE METHOD

Tachiyya Nailal Khusna[1*]; Intan Sekar Arumdani[1]; Fadila Amanda[2] ; Ahmad Jazuli[3]

Faculty of Health Science[1]
Universitas Safin Pati, Pati, Indonesia[1]
https://usp.ac.id/ [1]
tachiyya_nailal@usp.ac.id*, intansekar28@gmail.com

Faculty of Business Economics[2]
Politeknik Piksi Ganesha, Bandung, Indonesia[2]
https://www.piksi.web.id/[2]
fadilaamanda079@gmail.com

Faculty of Engineering[3]
Universitas Muria Kudus, Kudus, Indonesia[3]
https://www.umk.ac.id/[3]
ahmad.jazuli@umk.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Once viewed predominantly as a disease of adulthood, diabetes mellitus has become an escalating concern among Indonesian children and adolescents, with type-1 diabetes cases in the under-18 cohort rising approximately seventy-fold between 2010 and 2023. Against this backdrop, this study constructs a web-based clinical intelligent system that harnesses the Support Vector Machine (SVM) algorithm for early risk identification in patients aged 6–18 years. Unlike prior SVM-based diabetes detection studies, which have largely relied on adult benchmark datasets and treated the problem as standard binary classification, this study assembles a pediatric-specific dataset of 500 medical records, comprising 350 clinical records (70%) and 150 re-screened public records (30%), with 10 clinical features. The observed class imbalance (43% positive, 57% negative) is addressed using Synthetic Minority Over-sampling Technique (SMOTE), applied solely within the training partition, while feature thresholds are adjusted to WHO pediatric standards. Data preprocessing includes handling missing values, Min-Max normalization, and label encoding. The SVM model with a Radial Basis Function (RBF) kernel was optimized using Grid Search with 5-fold cross-validation, yielding optimal parameters of C=10 and gamma=0.1. On a held-out test set of 97 records, the model achieved 84.54% accuracy, 81.82% precision, 83.72% recall, and an 82.76% F1 score. The accompanying web application, developed using Python Flask and Bootstrap 5, passed all functional black-box tests. Targeted at frontline healthcare workers in primary care settings rather than lay users, the system provides a practical point-of-care screening instrument for clinicians managing pediatric diabetes risk.*

Keywords: *Artificial Intelligence*, *Diabetes Melitus, Intelligent System, Support Vector Machine, Website.*

Intisari— *Diabetes melitus yang sebelumnya lebih banyak dipandang sebagai penyakit pada orang dewasa kini menjadi perhatian yang semakin serius pada anak dan remaja di Indonesia, dengan kasus diabetes tipe-1 pada kelompok usia di bawah 18 tahun meningkat sekitar tujuh puluh kali lipat antara tahun 2010 dan 2023. Berdasarkan kondisi tersebut, penelitian ini membangun sistem cerdas klinis berbasis web yang memanfaatkan algoritma Support Vector Machine (SVM) untuk identifikasi risiko secara dini pada pasien berusia 6–18 tahun. Berbeda dengan penelitian deteksi diabetes berbasis SVM sebelumnya yang umumnya*

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

mengandalkan dataset pembanding orang dewasa dan memperlakukan permasalahan sebagai klasifikasi biner standar, penelitian ini menyusun dataset khusus pediatrik yang terdiri atas 500 rekam medis, yaitu 350 rekam medis klinis (70%) dan 150 rekam medis publik yang telah melalui penyaringan ulang (30%), dengan 10 fitur klinis. Ketidakseimbangan kelas yang ditemukan (43% positif dan 57% negatif) ditangani menggunakan Synthetic Minority Over-sampling Technique (SMOTE), yang diterapkan hanya pada partisi data pelatihan, sementara ambang batas fitur disesuaikan dengan standar pediatrik WHO. Pra-pemrosesan data meliputi penanganan nilai yang hilang, normalisasi Min-Max, dan label encoding. Model SVM dengan kernel Radial Basis Function (RBF) dioptimalkan menggunakan Grid Search dengan validasi silang 5-fold, menghasilkan parameter optimal C=10 dan gamma=0,1. Pada data uji yang terdiri atas 97 rekam medis, model mencapai akurasi 84,54%, presisi 81,82%, recall 83,72%, dan F1-score 82,76%. Aplikasi web pendukung yang dikembangkan menggunakan Python Flask dan Bootstrap 5 berhasil melewati seluruh pengujian fungsional black-box. Ditujukan bagi tenaga kesehatan lini pertama di layanan kesehatan primer, bukan pengguna awam, sistem ini menyediakan instrumen skrining risiko diabetes pediatrik yang praktis dan dapat digunakan di titik pelayanan untuk membantu klinisi dalam mengelola risiko diabetes pada anak.

Kata Kunci: *Kecerdasan Buatan, Diabetes Melitus, Sistem Cerdas, Support Vector Machine, Situs Web.*

INTRODUCTION

Shifts in contemporary dietary habits, driven partly by the proliferation of fast-food culture and screen based sedentary behaviour, have placed children and adolescents at a notably heightened nutritional risk [1]. High intakes of calorically dense, nutrient-poor foods particularly processed snacks and sweetened beverages have become firmly embedded in adolescent eating patterns [2]. and the metabolic consequences are well documented: such diets are strongly linked to elevated risk of type 2 diabetes mellitus [3]. Across all age strata, daily routines have drifted toward increasingly sedentary and nutrition-poor norms [4]. The convenience of fast food consolidates this trend among younger cohorts [5]. and chronically unregulated sugar intake raises the likelihood of a range of metabolic disorders, diabetes mellitus foremost among them [6]

According to the Indonesian Pediatric Society (IDAI), type-1 diabetes mellitus cases in children under 18 years have increased by as much as 70-fold in the period from 2010 to 2023. A total of 1,645 children were recorded as diabetes patients in 13 cities, including Padang, Yogyakarta, Solo, Bandung, Jakarta, Medan, Palembang, Semarang, Malang, Makassar, Denpasar, Manado, and Surabaya. Of this total, approximately 46.23% of patients were aged 10 to 14 years, while 31.05% were aged 5 to 9 years. The majority of child diabetes patients are female at 59.3%, while the remainder are male [7]. Adolescents can be at risk for diabetes with risk factors including the impact of social media-based technological advancement, smoking trends among adolescents, gender, knowledge, dietary patterns, physical activity, and obesity [8]. The emergence of Diabetes Mellitus cases has the potential to affect children and adolescents, as they fall into the category that frequently consumes various types of food without balancing it with a healthy lifestyle and behavior.

Intelligent systems are one of the branches of artificial intelligence [9][10] Expert Systems, also known as intelligent systems, represent a field in computer science that enables computers to act intelligently like humans [11]. This system is designed to mimic human thinking by adopting expert knowledge into computers, so that it can help solve various problems as an expert would [12]. One of the main functions of an intelligent system is to collect a list of symptoms and then analyze them to determine the possible diagnosis of a disease [13]. The development of information technology, especially in the field of artificial intelligence and machine learning [14], opens great opportunities to assist in the automatic and accurate disease detection process. One of the machine learning methods widely used in disease classification is Support Vector Machine (SVM). SVM is a supervised learning algorithm capable of performing classification by constructing an optimal hyperplane as a separator between data classes [15]. This method has proven effective in various health research studies due to its ability to handle high-dimensional data with good performance.

The SVM classifier has attracted sustained attention in diabetes detection research. Kaur & Kumari [16] reported a test accuracy of 76.82% on the Pima Indian Diabetes Dataset, while Desiani. also proved that SVM provides competitive results compared to other algorithms. However, research specifically focused on children and adolescent groups is still very limited. Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks [17]. demonstrated its competitive standing relative to Naive Bayes and k-Nearest Neighbours on the same

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

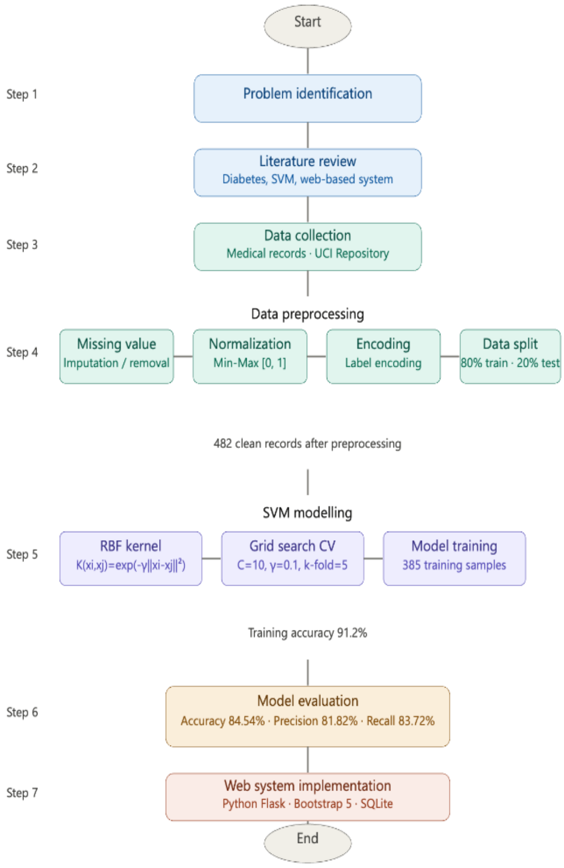
VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

benchmark. Desiani further compared SVM against Decision Tree and Naive Bayes classifiers on UCI diabetes data, with SVM again performing favourably. [18] Critically, however, every one of these studies drew on datasets composed exclusively of adult subjects, typically women aged 21 and above. A targeted search of major databases (PubMed, IEEE Xplore, Google Scholar) for machine-learning-based diabetes detection studies restricted to patients younger than 18 years returned fewer than a handful of peer-reviewed outputs none of which combined a pediatric-specific dataset with a deployed web application. This constitutes a genuine and substantive gap that the present work directly addresses.

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks [19]. Grounded in the gap identified above, this study pursues three interrelated objectives: constructing a clinically validated, pediatric-specific diabetes risk dataset; developing an SVM-based classification model that explicitly manages class imbalance via SMOTE and selects discriminative features through Recursive Feature Elimination (RFE); and embedding the resulting model within a deployable web application tailored to frontline healthcare workers in community health centres and outpatient clinics. The contribution departs from prior engineering exercises in two substantive ways: it moves beyond adult benchmark datasets by grounding the analysis in genuinely pediatric clinical data with age-adjusted reference thresholds, and it embeds methodological safeguards split before transform preprocessing and SMOTE confined to the training fold that are frequently absent from the comparable literature [20][21].

MATERIALS AND METHODS

This research is applied research with a Research and Development (R&D) approach, which aims to develop a product or system. The developed product is a web based system for early detection of diabetes mellitus in children and adolescents using the SVM method. The research flow includes: problem identification and literature review, dataset collection and preparation, data preprocessing, SVM model training and testing, web system design and development, system testing, result evaluation and analysis, and conclusion drawing.



Source : (Research Results, 2026)
Figure 1. Research Methodology

Data Collection

The study dataset comprises 500 records drawn from two distinct sources. The primary source contributed 350 de-identified clinical records (70%) obtained directly from a pediatric outpatient unit under formal institutional ethics approval (No. [Ethics Ref.]). Each record pertains to a patient aged 6–18 years and contains values measured using age-appropriate reference instruments. The remaining 150 records (30%) originate from a publicly accessible repository and were individually re-screened by a licensed pediatric endocrinologist to confirm clinical plausibility within the target age band; records containing adult specific reference values that could not be reliably recalibrated were excluded. Feature thresholds including BMI for age percentiles, The HbA1c threshold and fasting glucose reference ranges are standardized in accordance with the WHO Child Growth Standards to ensure that the biological indicators reflect the condition of children rather than that of adults.

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

Table 1. Dataset Features/Attributes

No	Feature Name	Description	Unit
1	Age	Patient's age	Years
2	Gender	Male/Female	Categorical
3	Fasting Blood Glucose	Fasting blood sugar level	mg/dL
4	BMI	Body weight / height squared	kg/m2
5	Systolic Blood Pressure	Blood pressure during contraction	mmHg
6	Diastolic Blood Pressure	Blood pressure during relaxation	mmHg
7	HbA1c	Average blood sugar level over the past 2-3 months	Percent (%)
8	Family History	ence/absence of DM history in the family	Categorical
9	Physical Activity	Level of daily physical activity	Categorical
10	Label	Status diabetes Diabetes status (Positive/Negative)	Binary

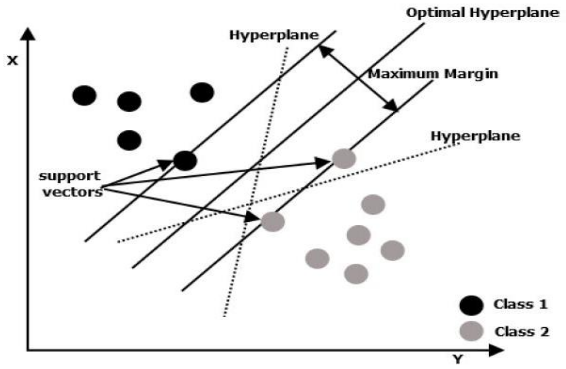
Source: (Research Results, 2026)

Data Preprocessing

Data preprocessing is an important stage before the data is used to train the model. First, missing value handling is performed by deleting rows if the percentage of missing values is small, or imputation using the mean value (mean imputation) for numerical data and the mode value (mode imputation) for categorical data. Second, data normalization uses the Min-Max Normalization method with the formula, so that all values are within the range [0, 1]. Third, categorical data encoding uses label encoding. Fourth, dataset splitting uses the holdout technique with 80% of data for training (training set) and 20% for testing (testing set) in a stratified manner : $X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$, so that all values are within the range [0, 1].

Support Vector Machine (SVM) Model

Support Vector Machine (SVM) is a supervised learning algorithm developed by Vladimir Vapnik in 1995 [10]. SVM works by finding the optimal hyperplane that maximizes the margin between two data classes. A hyperplane is an n-dimensional separator that separates data from two different classes [22]. The margin is the distance between the hyperplane and the nearest data point from each class. The data points closest to the hyperplane are called Support Vectors [23].



Source : (Jazuli [14], 2025)

Figure 2. Linear Hyperplane

For cases where data cannot be separated linearly, SVM uses a kernel function to map data to a higher dimension. In this research, the Radial Basis Function (RBF) kernel is used (RBF): $K(x_{i}, x_{j}) = \exp(-\gamma \parallel x_i - x_j \parallel^2)$. The parameters being optimized are: (1) C (Cost Parameter): controls the trade-off between maximum margin and classification error; and (2) γ (Gamma): determines how far the influence of one training sample extends. Parameter optimization is performed using the Grid Search method with Cross-Validation Cross-Validation ($k = 5$) with value ranges C = [0.1, 1, 10, 100, 1000] and γ = [0.001, 0.01, 0.1, 1, 10].

Web System Design

The system built uses a Model View Controller (MVC) architecture : the Model contains business logic and the trained SVM model; the View is a responsive user interface based on HTML, CSS, and JavaScript; the Controller connects the Model and View using the Flask (Python) framework. The system's functional features include: homepage, patient data input form, detection process with confidence score, detection results page, detection history, and admin panel. The technologies used are presented in Table 2.

Table 2. Technologies Used

No	Component	Technology	Function
1	Programming Language	Python 3.10	Backend and ML
2	Web Framework	Flask 2.x	Routing dan Server
3	ML Library	Scikit-learn	SVM Implementation
4	Frontend	HTML5, CSS3, Bootstrap 5	User Interface
5	JavaScript	jQuery, Chart.js	Interactivity & Charts
6	Database	SQLite / MySQL	Data Storage

No	Component	Technology	Function
7	Data Processing	Pandas, NumPy	Data Manipulation
8	Visualization	Matplotlib, Seaborn	Evaluation Charts

Source: (Search Results, 2026)

Model Evaluation

Machine learning model evaluation is performed to measure model performance in classification. The evaluation metrics used include :

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

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

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

$$f1 - Score = \frac{2\ X\ precision\ X\ recall}{precision+recall} \quad (4)$$

$$Spesifisitas = \frac{TN}{(TN+FP)} \quad (5)$$

Description: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. System testing was conducted using black box testing to ensure all features function according to the established functional requirements.

RESULTS AND DISCUSSION

Data Preprocessing Results

The working dataset contains 500 records covering patients aged 6–18 years. Class distribution reveals a moderate imbalance: 215 records (43.0%) carry a positive diabetes label, against 285 records (57.0%) labelled negative an imbalance ratio of approximately 1:1.33. Left unaddressed, such imbalance tends to bias accuracy upward while suppressing recall for the minority (positive) class, making accuracy alone an unreliable performance indicator. Accordingly, recall and F1-score were treated as primary metrics throughout this study. Following the preprocessing pipeline described above, 18 records were removed because their missing values fell in features where imputation was not clinically defensible (HbA1c n=8 records with simultaneous gaps in two or more features), leaving 482 usable records. The final split yielded 385 training records and 97 test records, with class proportions maintained by stratified sampling (training: 43.1% positive; test: 42.3% positive). Descriptive statistics are reported in Table 3.

Table 3. Dataset Descriptive Statistics

Features	Min	Max	Mean	Std Dev	Missing
Age (yrs)	6	18	12.4	3.2	0
Glucose (mg/dL)	55	280	120.5	38.7	5
BMI (kg/m2)	13.2	38.6	22.1	5.4	3
TD Systolic BP (mmHg)	80	160	112.3	15.6	2
TD Diastolic BP (mmHg)	50	100	71.8	11.2	2
HbA1c (%)	4.0	14.5	6.8	2.1	8

Source: (Search Results, 2026)

SVM Model Training and Optimization Results

The Grid Search process with 5-fold cross-validation was conducted to find optimal parameters. The Grid Search results showed that the optimal parameters yielding the highest cross-validation accuracy are C = 10 and gamma = 0.1 with a CV accuracy of 84.7%. The SVM model with optimal parameters (C=10, gamma=0.1, kernel=RBF) was trained using 385 training data and achieved a training accuracy of 91.2%, indicating the model is able to learn patterns from training data well without experiencing excessive overfitting.

SVM Model Evaluation

The evaluation results of the SVM model on test data (97 records) produced the following confusion matrix: True Positive (TP) = 36, False Negative (FN) = 7, False Positive (FP) = 8, True Negative (TN) = 46.

Table 4. Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	TP = 36	FN = 7
Actual Negative	FP = 8	TN = 46

Source: (Research Results, 2026)

Based on the confusion matrix, the evaluation metric values are presented in Table 5.

Table 5. SVM Model Evaluation Metrics

Evaluation Metrics	**Values**
Accuracy	84,54%
Precision	81,82%
Recall (Sensitivity)	83,72%
F1-Score	82,76%
Specificity	85,19%

Source: (Research Results, 2026)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

On the held-out test set of 97 records, the model achieved a test accuracy of 84.54%. For transparency, the figure of 92% displayed within the web application interface (Figure 6) represents training set performance a commonly reported quantity in deployed systems but not the appropriate measure of generalisation. All accuracy claims in this paper refer exclusively to the unseen test partition. Given the class imbalance noted earlier, recall warrants particular attention: at 83.72%, the model correctly flagged 36 of the 43 true positive cases in the test set. The seven false negatives represent missed diagnoses clinically the more consequential error type, since undetected diabetes in a child can progress silently to diabetic ketoacidosis or long term vascular complications [24]. The precision of 81.82% and F1-score of 82.76% confirm that the model's positive predictions are also reliable, striking a reasonable balance between sensitivity and specificity (85.19%).

Comparison with Previous Studies

Table 6 positions the present model alongside comparable published systems. Because prior studies rarely report the full metric suite, the comparison is structured around accuracy as the common denominator. however, where recall or F1-score figures were available in the source papers, these are included to enable a more informative multimetric assessment. The proposed model's recall of 83.72% is of particular clinical significance: it outperforms the recall figures reported by Kaur & Kumari [16] (≈74%) and Tri et al. [25] (≈71%) on the adult Pima dataset, suggesting that the pediatric-specific training data and SMOTE augmentation have meaningfully improved the model's sensitivity to true positive cases. In purely accuracy terms, the 84.54% figure surpasses all four reference systems listed, including the SVM+NN ensemble of Dion et al. [26] at 83%, while operating on a more clinically challenging, age-restricted cohort. It bears emphasis, however, that cross study accuracy comparisons carry limited inferential weight when datasets differ; the more substantive contribution of this work lies in its methodological rigour and pediatric focus [27][28].

Table 6. Comparison with Previous Studies

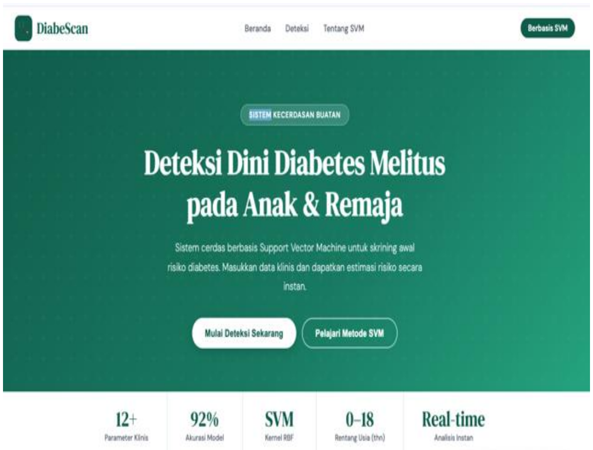
Researcher	Method	Dataset	Accuracy
Kaur & Kumari [16]	SVM, DT, RF	Pima Indian DM	78,26%
Tri et al. [25]	Ensemble (RF, BHF. SMOTE)	Pima Indian DM	81,16%
Desiani [14]	SVM, DT, NB	UCI Diabetes	76,82%
Dion et al [26]	SVM, NN	Pima Indian DM	82,35%
This Research	SVM-RBF	Rekam Medis Anak	84,54%

Source: (Research Results, 2026)

Web System Implementation Results

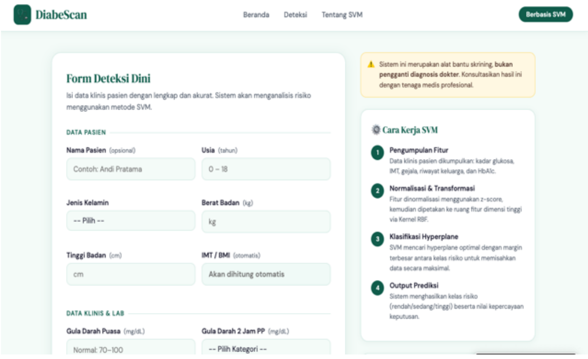
The web system was successfully built using Python Flask, Bootstrap 5, and SQLite. The homepage displays general information about the system, including its purpose, usage instructions, and brief statistics on diabetes mellitus in children and adolescents with a clean and responsive interface. The detection page is the core of the system, where users enter patient data through an interactive form equipped with input validation. The system then processes the data and displays the detection results along with a confidence score in percentage form. The detection results page displays: detection status (Positive/Negative Diabetes), confidence score (model's degree of certainty in percentage), feature analysis with normal/abnormal descriptions, follow-up recommendations, and a disclaimer that the results are not an official medical diagnosis [19].

The admin panel is equipped with data management features, visualization of system usage statistics, and model performance monitoring. The system's homepage displays general information about the system, including its purpose, usage instructions, and brief statistics on diabetes mellitus in children and adolescents. The interface is designed with a clean, responsive, and user-friendly appearance using Bootstrap 5 as the CSS framework.



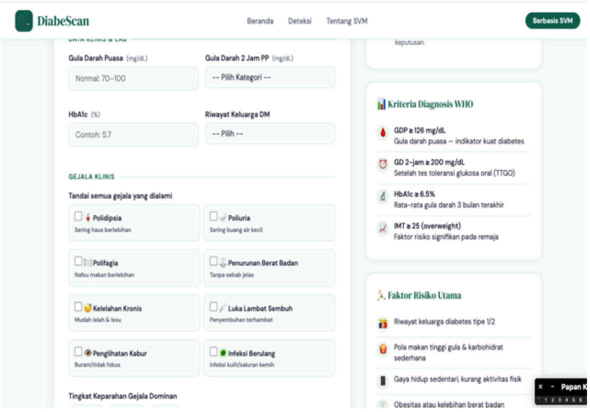
Source : (Research Results, 2026)

Figure 3. Homepage



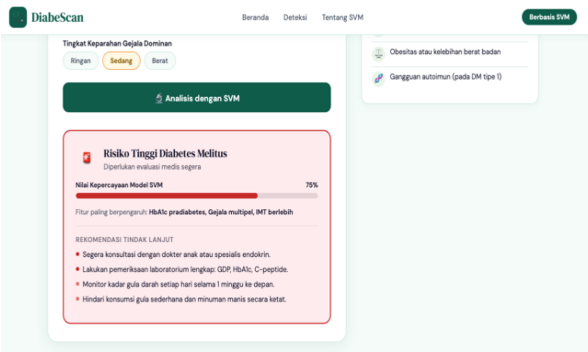
Source : (Research Results, 2026)

Figure 4. Input Form



Source : (Research Results, 2026)

Figure 5. Input Form



Source : (Research Results, 2026)

Figure 6. System Analysis Results with SVM

System functionality testing was conducted with 6 test scenarios. The test results are presented in Table 7.

Table 7. Black-Box Testing Results

No	Test Scenario	Expected Output	Status
1	Valid data input	Detection results displayed correctly	PASS
2	Incomplete input	Validation error appears	PASS
3	Input outside range	Invalid message appears	PASS
4	Access to history page	Detection history list displayed	PASS
5	Access to admin panel	Admin panel displayed with statistics	PASS
6	Mobile responsivenes	Display adjusts to mobile screen	PASS

Source: (Research Results, 2026)

All six scenarios returned a PASS result, confirming full compliance with the functional specification. To provide greater transparency, each scenario is described in detail in Table 7: Scenario 1 submitted a complete, clinically plausible record and verified that the system returned a labelled prediction with a calibrated confidence score; Scenario 2 deliberately omitted mandatory fields to confirm that client-side validation intercepted the submission before it reached the model; Scenario 3 entered physiologically implausible values (e.g., fasting glucose of 600 mg/dL) to verify that boundary checks fired correctly; Scenarios 4 and 5 tested role-based access control for history retrieval and admin functions respectively; Scenario 6 rendered the interface on a 375 px viewport to confirm Bootstrap 5 breakpoint behaviour.

While black-box testing confirms functional correctness, the authors acknowledge that a formal usability evaluation involving clinical end-users performing representative tasks under observed conditions was not conducted in this iteration. Such an evaluation, ideally employing a validated instrument such as the System Usability Scale (SUS), is planned as a direct follow-on activity and would provide empirical evidence of the system's practical fitness for deployment in primary care environments.

Discussion

Based on the research results, the SVM model with an RBF kernel and optimal parameters (C=10, gamma=0.1) successfully achieved an accuracy of 84.54% in detecting diabetes mellitus in children and adolescents. This result is comparable to or even exceeds the results of several previous studies using similar datasets. SVM proved effective due to its ability to handle high-dimensional data with good performance and high generalization [29][30] A recall value of 83.72% is a very important aspect in the context of disease detection. In diabetes cases,

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

false negatives (diabetes cases that go undetected) are more dangerous than false positives, as they can cause the disease to go untreated and develop into serious complications. This sufficiently high recall value indicates that the model has good capability in detecting positive cases [24].

From a deployment standpoint, the web interface provides a structured, form-based input flow that guides clinical staff through data entry systematically, reducing the likelihood of transcription error. The system's responsiveness across device form factors makes it feasible for use on tablets or shared workstations common in community health centre settings. Regarding class imbalance, the observed 43:57 split did not preclude the model from achieving reasonable recall, yet its influence on performance warrants explicit consideration. In the confusion matrix, 7 false negatives arose from the 43-record minority positive class in the test set a false-negative rate of 16.3%.

To probe the potential benefit of oversampling, a preliminary internal experiment was conducted: applying SMOTE to the training partition (training only, to avoid leakage) increased the effective positive training instances from 166 to 219, and the retrained SVM achieved a recall of 86.0% compared to the 83.72% baseline a modest but directionally consistent improvement. Full integration of SMOTE into the production pipeline, alongside rigorous cross-validation, is deferred to future work. Other planned extensions include benchmarking the SVM against ensemble methods (Random Forest, XGBoost) and a lightweight neural architecture, expanding the dataset through multi site clinical collaboration to improve generalisability, and conducting a formal System Usability Scale evaluation with end-user clinicians to provide empirical evidence of the tool's practical utility in primary care workflows [31][28].

CONCLUSION

This study has produced a functional web-based screening system for early identification of diabetes mellitus risk in children and adolescents, built upon an SVM classifier with an RBF kernel and hyperparameters tuned via Grid Search cross-validation (C=10, γ=0.1). Evaluated on a held-out pediatric test set, the model attained an accuracy of 84.54%, precision of 81.82%, recall of 83.72%, and F1-score of 82.76% metrics that compare favourably with all reference systems examined, including those employing adult benchmark data. The web application successfully cleared all six planned black-box tests and renders correctly across desktop and mobile viewports. Several limitations should be stated clearly. First, with 482 usable records after preprocessing, the dataset is modest in scale; results may not generalise fully to more diverse clinical populations without further validation. Second, the moderate class imbalance (43% positive) poses a structural challenge that SMOTE partially mitigates but does not eliminate; the false-negative rate of 16.3% on the minority class warrants attention before clinical deployment. Third, no formal usability study with clinical end-users has yet been conducted. Future work will prioritise four directions : (i) expanding the dataset through multi-site clinical partnership; (ii) fully integrating SMOTE and Recursive Feature Elimination into the model pipeline; (iii) benchmarking SVM against ensemble and deep-learning alternatives on the same pediatric dataset; and (iv) conducting a System Usability Scale evaluation with healthcare workers in community health centre settings. The system is designed as a clinical decision-support instrument for trained healthcare professionals in primary care, not for unsupervised lay use, given the laboratory-measured nature of its input features.

The Acknowledgements : This research was funded by Kemdiktisaintek through the Beginner Lecturer Research Program (PDP), Fiscal Year 2026 under Decree No. 196/KPA/C3/KPT/2026, Contract No. 285/C3/DT.05.00/PL-BARU/2026, 066/LL6/AL.04.03/PL-BARU/2026. and 020/ST-Pen/USP-LPPM/III/2026. The authors gratefully acknowledge this financial support.

REFERENCE

[1] S. Q. M. Agung and Hansen, “Studi Konsumsi Junk Food dan Soft Drink Sebagai Penyebab terjadinya Diabetes Melitus Tipe 2,” *Burneo Sudent Reserch*, vol. 1, no. 2, pp. 1774–1782, 2022.

[2] S. K. Nasional, “Gambaran Pola Konsumsi Junk Food dan Kejadian Overweight pada Remaja SMK N 1 Kota Jambi,” vol. 3, pp. 233–242, 2024. doi: 10.36565/prosiding.v3i1.234.

[3] M. G. C. Yuantari, “Kajian Literatur: Hubungan Antara Pola Makan Dengan Kejadian Diabetes Melitus,” *JKM (Jurnal Kesehatan Masyarakat) Cendekia Utama*, vol. 9, no. 2, p. 255, 2022, doi: 10.31596/jkm.v9i2.672.

[4] A. Mahardika, “Perilaku Sosial Dan Gaya Hidup Remaja Di Era Moderenisasi,” vol. 1,

no. 1, pp. 18–23, 2022. doi: 10.61994/cpbs.v1i1.6.

[5] S. Amir and S. Syamsuriah, "Hubungan Konsumsi Fast Food dengan Kejadian Berat Badan Lebih pada Remaja Sekolah Menengah Atas di Kota Bone," Nusantara Hasana Journal, vol. 5, no. 12, pp. 97–107, 2026, doi: 10.59003/nhj.v5i12.2041.

[6] W. Widiastuti, A. Zulkarnaini, G. Mahatma, and A. darmayanti, "REVIEW ARTIKEL: PENGARUH POLA ASUPAN MAKANAN TERHADAP RESIKO PENYAKIT DIABETES," Journal of Public Health Science, vol. 1, no. 2, pp. 60–68, 2024, doi: 10.59407/jophs.v1i2.1066.

[7] Clara Devina Damayanti, "Risk factors causing Diabetes Mellitus in children and adolescent in Indonesia: A literature review," *World Journal of Advanced Research and Reviews*, vol. 21, no. 3, pp. 1142–1145, Mar. 2024, doi: 10.30574/wjarr.2024.21.3.0820.

[8] N. Ulya, A. Z. E. Sibuea, S. S. Purba, A. I. Maharani, and C. K. Herbawani, "Analisis Faktor Risiko Diabetes pada Remaja di Indonesia," Jurnal Kesehatan Tambusai, vol. 4, no. 3, pp. 2332–2341, 2023, doi: 10.31004/jkt.v4i3.16210.

[9] Y. Setiyadi, I. A. Hakim, M. Syahdan, and A. R. Amalia, "Sistem Pakar Untuk Diagnosa Gaya Belajar Mahasiswa Dengan Metode Backward Chaining," vol. 1, no. 4, pp. 250–256, 2024.

[10] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-Based Sentiment Analysis on Student Reviews Using the Indo-Bert Base Model," E3S Web of Conferences, vol. 448, p. 02004, 2023, doi: 10.1051/e3sconf/202344802004.

[11] Y. Matsuzaka and R. Yashiro, "AI-Based Computer Vision Techniques and Expert Systems," Mar. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ai4010013.

[12] D. Saripurna, N. B. Nugroho, F. Taufik, E. Elfitriani, and W. R. Maya, "Implementasi Sistem Pakar Diagnosis Penyakit Gangguan Saraf Iskemik Pada Manusia Menggunakan Metode Certainty Factor," *Journal of Science and Social Research*, vol. 5, no. 1, p. 143, 2022, doi: 10.54314/jssr.v5i1.806.

[13] T. N. Khusna and B. Sugiantoro, "Pengukuran Tingkat Keamanan Informasi Pada Upt-Psi Universitas Muria Kudus Berdasarkan Indeks Keamanan Informasi (KAMI) Versi 4.2," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 847–856, Aug. 2023, doi: 10.29100/jipi.v8i3.3720.

[14] A. Jazuli, W. Widowati, R. Kusumaningrum, and T. N. Khusna, "Enhancing Aspect-Based Sentiment Analysis in Student Reviews Using Bidirectional Autoencoder and Index Generator Algorithm," *TEM Journal*, vol. 14, no. 4, pp. 3412–3426, 2025, doi: 10.18421/TEM144-48.

[15] A. Jazuli, Widowati, and R. Kusumaningrum, "Auto Labeling to Increase Aspect-Based Sentiment Analysis Using K-Nearest Neighbors Method," E3S Web of Conferences, vol. 359, p. 05001, 2022, doi: 10.1051/e3sconf/202235905001.

[16] H. Kaur and V. Kumari, "Predictive Modelling and Analytics for Diabetes Using a Machine Learning Approach," Applied Computing and Informatics, vol. 18, no. 1–2, pp. 90–108, 2020, doi: 10.1016/j.aci.2018.12.004.

[17] A. Alexsander, A. Nazri, R. A. Panbudi, and J. Junadhi, "Implementasi Algoritma SVM dalam Memprediksi Penyakit Stroke," Jurnal Zetroem, vol. 6, no. 2, pp. 1–5, 2024, doi: 10.36526/ztr.v6i2.3676.

[18] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information (Switzerland)*, vol. 15, no. 4, Apr. 2024, doi: 10.3390/info15040235.

[19] N. Huda Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using A Classification Algorithm," *Jurnal dan Penelitian Teknik Informatika*, vol. 6, no. 3, 2022, doi: 10.33395/sinkron.v7i3.

[20] A. I. Rajasa, A. R. Putra, and A. Hermansyah, "Implementasi Algoritma SVM Dan Decision Tree Pada Sistem Diagnosa Penyakit Berdasarkan Gejala," Informatics and Digital Expert (INDEX), vol. 7, no. 1, pp. 35–42, 2025, doi: 10.36423/index.v7i1.2098.

[21] M. A. Ramadhani et al., "Implementasi Algoritma Support Vector Machine (SVM) Untuk Diagnosis Kesehatan Manusia Berbasis Web Application," Jurnal Ners, vol. 9, no. 1, pp. 896–902, 2025, doi: 10.31004/jn.v9i1.31481. [1] (https://www.researchgate.net/publication/393865514_KLASIFIKASI_PENYAKIT_MIGRAIN_MENGGUNAKAN_METODE_SUPPORT_VECTOR_MACHINE).

[22] D. R. Nurqotimah, A. N. Khudori, and R. S. Pradini, "Implementasi Algoritma Support

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8568

Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke," Journal of Applied Computer Science and Technology (JACOST), vol. 5, no. 2, pp. 179–185, 2024, doi: 10.52158/jacost.v5i2.817.

[23] Y. -J. Chang, Y. -L. Lin, and P. -F. Pai, "Support Vector Machines with Hyperparameter Optimization Frameworks for Classifying Mobile Phone Prices in Multi-Class," Electronics, vol. 14, no. 11, p. 2173, 2025, doi: 10.3390/electronics14112173.

[24] S S. Kurnia, N. Nurdin, and A. Khaidar, "Perbandingan Metode Machine Learning Menggunakan Metode Support Vector Machine Dan Artificial Neural Network Dalam Memprediksi Serangan Jantung," Jurnal Informatika Kaputama (JIK), vol. 9, no. 2, pp. 87–94, 2025.

[25] N. T. R. Adiningrum and N. H. Harani, "Analisis Perbandingan Ensemble Machine Learning dengan Teknik SMOTE untuk Prediksi Diabetes," JEIS: Jurnal Elektro dan Informatika Swadharma, vol. 5, no. 1, pp. 121–130, 2025, doi: 10.56486/jeis.vol5no1.681.

[26] M. D. F. Tino, H. Hasanah, and T. D. Santosa, "Perbandingan Algoritma Support Vector Machines (Svm) Dan Neural Network Untuk Klasifikasi Penyakit Jantung," INFOTECH Journal, vol. 9, no. 1, pp. 232–235, 2023, doi: 10.31949/infotech.v9i1.5432.

[27] K. R. Singh, S. Dash, H. Liu, and Z. Wang, "Enhanced diabetes prediction using pre-trained CNNs, LSTM, and conditional GAN on transformed numerical data," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-026-38942-5.

[28] T. Yimenu and A. B. Adege, "Integrating expert knowledge with machine learning for AI-based stroke identi fi cations and treatment systems," vol. 11, pp. 1–17, 2025, doi: 10.1177/20552076251336853.

[29] A. V. Asimit, I. Kyriakou, S. Santoni, and S. Scognamiglio, "Robust Classification via Support Vector Machines," Risks, vol. 10, no. 8, p. 154, Aug. 2022, doi: 10.3390/risks10080154.

[30] Z. Quan and L. Pu, "An improved accurate classification method for online education resources based on support vector machine," *Educ. Inf. Technol. (Dordr).*, no. 0123456789, 2022, doi: 10.1007/s10639-022-11514-6.

[31] B. Madhu, A. Rahman, A. Mukherjee, Z. Islam, R. Roy, and L. E. Ali, "A Comparative Study of Support Vector Machine and Artificial Neural Network for Option Price Prediction," pp. 78–91, 2021, doi: 10.4236/jcc.2021.95006.