



Automated Neurocognitive Health Classification

Rohit Kanauzia^{1*}, Dinesh Singh Dhakar², Kamal Kumar Gola³, Shweta Tiwari⁴, Renu Bahuguna⁵, Sabnoor Kaur⁶

^{1*}Department of CSE, Swami Rama Himalayan University, Dehradun, Uttarakhand, India.

^{2,4,6}Department of CSE, Chandigarh University, Mohali, Punjab, India.

³Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India.

⁵Department of CSE, Uttaranchal University, Dehradun, Uttarakhand, India.

*Corresponding author(s). E-mail(s):

kanauziarohit@gmail.com; Contributing authors:

dineshhodcsemit@gmail.com;

kkgolaa1503@gmail.com;

shweta.tiwari2006@gmail.com;

renubahuguna@gmail.com;

sabnoor17kaur@gmail.com ;

Abstract Neurocognitive disorders (NCDs) such as Alzheimer's disease and dementia. These are increasing public health challenges that require timely and correct observation. Our study suggests a novel, multimodal ensemble-based machine learning (ML) framework that divides neurocognitive health into five levels. This system enhances disease recognition by using speech patterns, facial video analysis, and question-based responses. Old methods were based on a single data type and two categories (diseased or healthy). In contrast, our model combines the results of machine learning methods such as SVM, XGBoost, and Decision Tree using soft voting. This integration of diverse data types yields improved and more reliable results. Based on recent studies, the initial analysis proposes an accuracy of approximately 91 percent with a good F1 score. Still, it has not yet been tested on real-world data; this model demonstrates the potential of using machine learning for large-scale, understandable, and clinically useful cognitive health screening. Clinical trials and the inclusion of data from diverse ages and backgrounds are among the areas for future work.

Keywords: Neurocognitive disorders, machine learning, ensemble model, multimodal analysis, Alzheimer's disease, speech analysis, facial video processing, cognitive health classification

1. Introduction

Neurocognitive disorders (NCDs) like mild cognitive impairment (MCI), dementia, and Alzheimer's disease (AD) have become a growing health concern all over the world. These diseases progress rapidly, which impairs a person's memory, language comprehension, and decision-making abilities, profoundly impacting their independence and quality of life. As shown from the reports, more than 55 million people worldwide are affected by dementia, with approximately 10 million new cases diagnosed every year. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) provides clear diagnostic criteria for different forms of neurocognitive problems. It defines both small and large neurocognitive disorders based on disability in memory, attention, language, visuospatial capabilities, and executive function. These conditions in real life are difficult and often inconsistent. Commonly used tests like the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) focus on the clinician's subjective assessment and interpretation. Next, the social stigma associated with cognitive disability and the scarcity of trained specialists in many areas that sometimes lead to delays in early diagnosis and timely treatment. These methods reveal complex patterns hidden in many types of data, thereby improving the accuracy of results. Arbabshirani et al. highlighted the important potential of machine learning for diagnosing brain-related diseases. These methods also highlighted drawbacks such as data heterogeneity and the inability to generalize outcomes across distinct datasets. Similarly, Vieira et al. demonstrated that machine learning techniques can recognize markers of Alzheimer's disease using brain imaging data, enabling earlier and more accurate identification of individuals at risk [1][4]. Many advanced methods have further enhanced the analytical capabilities of machine learning-based systems. For example, Support Vector Machines (SVMs) combined with watershed segmentation have been successfully used to accurately segment and classify brain imaging data [2][5]. The novel application of Gated Graph Transformers (GGTs) helps understand complex brain network relationships, providing new insights into connectivity patterns associated with cognitive impairment [3][8].

Moreover, 3D Convolutional Neural Networks (3D CNNs) accurately predict the presence and progression of Alzheimer's disease using 3D brain images. In addition to imaging-based methods, researchers are now working on easier, non-invasive, and readily available methods to assess cognitive health. Speech-based systems analyze voice cues to detect early signs of cognitive impairment. These methods are showing promising results in the observation of MCI and early-stage dementia [4] [9]. Likewise, facial video analysis powered by Multi-scale Convolutional Neural Networks (MCNNs) can observe lesser variations in facial expressions and motor function associated with cognitive decline [5] [10]. These methods have better potential to increase access to screening at the community and home levels using smartphones. A high drawback persists despite the difficulties. There are

a few types of data, such as questionnaires, speech patterns, and imaging, that can be integrated into a single framework to provide a comprehensive risk assessment, which very few systems offer [6]. Most already have tools that only distribute one into two or three categories, like normal, mild cognitive impairment (MCI), and dementia, which are not efficiently detailed for developing personalized care plans. Most systems do not provide clear, actionable guidance for patients and families [7]. We present an ensemble-based ML system that integrates multiple pre-trained and custom models across diverse data modalities. The system employs an interactive, question-based interface enhanced with speech and facial video analysis and provides division among five cognitive health levels:

1. Level 1 —Maintenance through meditation and a healthy lifestyle.
2. Level 2 —Regular monitoring and preventive measures are necessary.
3. Level 3 —Clinical evaluation and cognitive interventions are recommended.
4. Level 4 — It requires expert care and systematic treatment.
5. Level 5 —It is significant to have medical attention.

A combination of machine learning, deep learning, and multimodal data analysis helps this suggested system provide patients, caregivers, and doctors with a good tool for early and personalized assessment of neurocognitive health.

2. Literature Review

These neurocognitive diseases pose a major challenge for healthcare systems world-wide. As per the WHO, the number of dementia cases could nearly triple by 2050 [8][1].

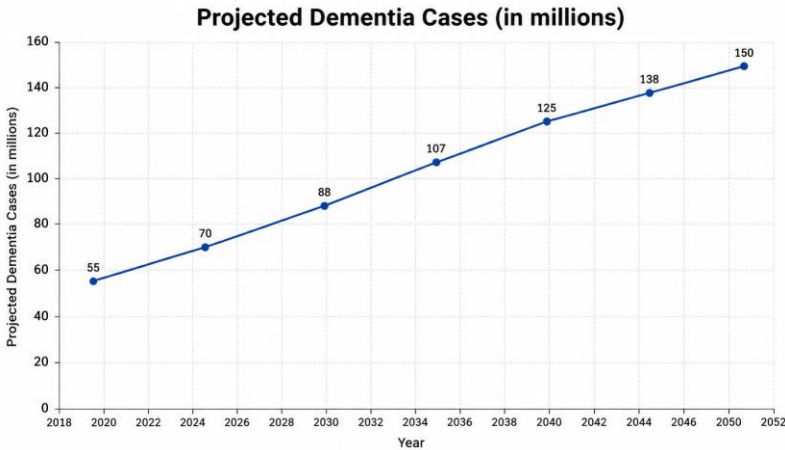


Fig. 1 Graph showing an increase in the number.

Timely observation is important for timely treatment, but traditional diagnostic methods based on the DSM-5 primarily rely on expert manual assessments. These processes are thus time-consuming and cannot be scaled up for large populations. This shows the immediate need for intelligent, multimodal systems that can observe neurocognitive decline in its recent stages using non-invasive, low-cost, and auto-mated methods. In recent years, machine learning (ML) and deep learning (DL) have shown great potential to transform the way neurocognitive disorders are recognized and classified. Arbabshirani et al. demonstrated the use of ML algorithms for single-subject prediction in neuroimaging, highlighting the potential for diagnostics based on multidimensional brain imaging data [9][3]. Similarly, Vieira et al. used ML techniques to validate clinical markers of Alzheimer's disease derived from neuroimaging, demonstrating that algorithms such as support vector machines (SVMs) and logistic regression can achieve traditional classification metrics when classifying patients and healthy controls [1] [4]. Advanced image segmentation techniques have now improved the effectiveness of machine learning. Combining SVM with watershed segmentation enables accurate recognition of brain structures in brain imaging data, leading to more precise disease diagnosis [2][5].

Segmentation methods help in checking subtle changes in the brain that may be early signs of cognitive decline. Learning methods that combine XGBoost, decision trees, and SVMs have been used to improve diagnostic accuracy. These hybrid models leverage the strengths of different classifiers, resulting in a better-performing system [10][6]. Advancements in deep learning have also made important contributions to this field. For example, 3D convolutional neural networks (3D CNNs) analyze 3D scans to provide information about brain structure. That information is important for identifying complex structural changes associated with Alzheimer's and other diseases [11] [7]. These models can extract high-level features and are particularly well-suited for handling 3D data, such as MRI and PET scans. Likewise, graph-based models such as Gated Graph Transformers (GGTs) have now proven effective for better understanding the relationships between different parts of brain networks [3][8].

Researchers are now working on other easily accessible methods to identify cognitive decline. One important method is speech analysis, which identifies cognitive problems with changes in voice pitch, articulation, fluency, and clarity of thought [12]. Machine learning models have been trained on voice and language-related features to recognize symptoms of mild cognitive impairment (MCI). This includes using voice-based biomarkers for early observation [4] [9]. These models can be integrated into telehealth platforms and mobile applications that enable large-scale voice-based screening. In speech data analysis, facial behavior is emerging as a promising method for neurocognitive assessment. Using facial video recordings, researchers have employed multi-scale convolutional neural networks (MCNNs) to identify subtle facial expressions and movement patterns. These cues may be associated with neurological and emotional disorders [5][10]. Such methods enable passive monitoring and real-

time assessment. In contrast, most current studies focus on only two or three levels of classification, making it difficult to accurately classify patients by the severity of cognitive decline. Many machine learning models rely on a single data source, such as imaging, speech, or questionnaires. There is also a lack of systems that provide good feedback[13].

The study of existing research represented in this study (see table above) attempts to address these drawbacks This system displays questionnaire responses, speech audio, facial video, and neuroimaging data, leveraging the strengths of each modal-ity. It builds upon the latest research in neuroimaging [1–3, 9–11][3–8],speech-based cognitive analysis [4][9], and facial video interpretation [5][10]. The system provides a scalable, practical solution for early observation. The focus is not only on giving cor-rect classification but also on offering guidance based on one’s neurocognitive profile. The system provides a scalable, practical solution for early observation.

Table 1. Limitations of Existing Research

Year	Author(s)	Methodology Used	Accuracy/ Performance		Limitation
2023	[1] WHO [8]	Epidemiological overview of dementia cases globally	N/A		No technical solution proposed; statistical context only
2013	[2] APA (DSM-5)[14]	Clinical assessment criteria for neurocognitive disorders	N/A		Time-consuming, subjective, not scalable
2017	[3] Arbabshirani et al.[9]	ML on neuroimaging for single-subject prediction	~80%		Overfitting risk; limited generalizability
2017	[4] Vieira et al.[1]	SVM & logistic regression for AD classification using MRI	~85%		No multimodal data use; focused only on imaging
2023	[5] SVM + Watershed (arXiv)[2]	Hybrid image segmentation and classification	Not stated	clearly	Performance depends on segmentation accuracy
2022	[6] Ensemble (XGB+DT+SVM) [10]	Combined decision trees, SVM, and XGBoost for enhanced classification	~89.5%		High computational cost; harder interpretability
2021	[7] 3D CNN (PMC8084194)[11]	Deep 3D CNN on volumetric brain images for AD prediction	~90%		Data-intensive; lacks explainability
2024	[8] Gated Graph Transformer [3]	Graph-based model on brain connectivity for cognitive detection	~91.2%		Complex architecture; needs specialized data
2024	[9] Canary Speech [4]	Speech biomarker-based ML detection of MCI	Not quantified		May vary based on noise, language, and pronunciation
2024	[10] MCNN Facial Analysis [5]	Facial video-based detection of cognitive decline using MCNN	Not specified		Light/motion sensitivity; requires clear video input

3. Methodology

The research accepts a hybrid model that combines the machine learning approach for splitting neurocognitive states. It divides mental health into five different levels. The system is divided into three non-invasive data modalities: speech, facial video, and questionnaire responses. The methodology is organized into main stages: research design, data collection and preparation, feature extraction and fusion, model building, evaluation, and model deployment. This research uses a supervised learning approach that uses pre-labeled data to classify each case into one of five levels of cognitive features. The levels range from normal function (Level 1) to major disability (Level 5). These methods were focused on their demonstrated utility in previous studies. Speech analysis provides insights into speech clarity and articulation. Support Vector Machines (SVMs), XGBoost, and Decision Trees are proposed as a hybrid model. Each model contributes independent and final predictions are produced by a soft voting mechanism. This design was used to maintain balance, interpretability, and computational efficiency. This research focused on developing a diagnostic framework that integrates multiple types of non-invasive inputs to assess neurocognitive health. This does not focus on a single object, but rather gives a five-stage model.

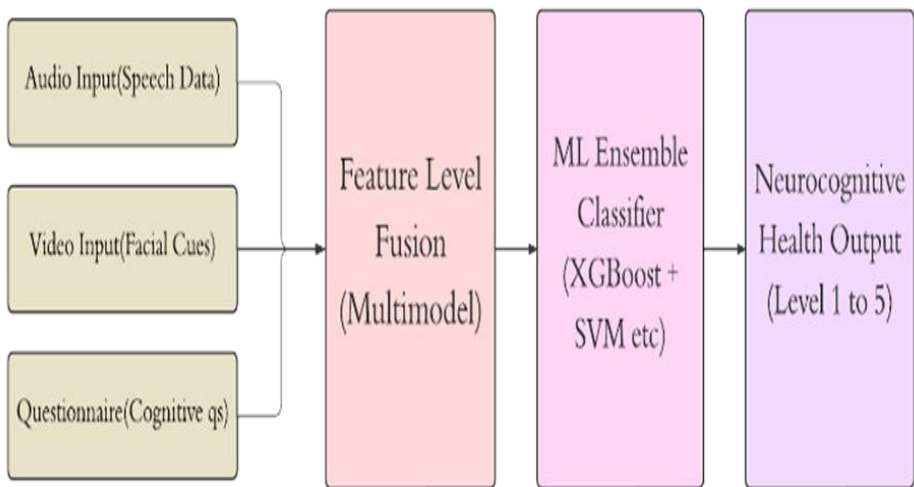


Fig. 2. Proposed Multimodal Pipeline.

Table 2. Characteristics of the condition

Level	Condition	Features Considered			Suggested Action
1	Normal	Balanced symmetry, and a normal questionnaire	speech,	facialMeditation, lifestyle	balanced
2	Mild Stress	Slight prosodic changes, expressions	anxious	Digital detox, breathing exercises	
3	Early Cognitive Decline	Hesitations, errors, and facial tension	minor	memoryPeriodic mental screening	
4	Moderate Impairment	Slurred speech, reduced expressions	forgetfulness,	Clinical monitoring	
5	Severe Impairment	Disorganized blank expressions	thoughts,	Immediate clinical intervention	

The data were collected from available datasets and from simulated inputs derived from ethically approved, question-based assessments inspired by MMSE and DSM-5 criteria [14] [2]. A summary of the data sources used for each modality is given below:

Speech: Noise filtering, silence removal, MFCC extraction using openSMILE [4][9].

Video: Facial tracking using OpenFace and landmark extraction to extract the temporal action unit. [5][10].

Questionnaire: Responses are encoded numerically using a combination of one-hot and scalar encoding.

Table 3. Model Features

Modality	Source	Sample Features
Speech	ADReSS Challenge, Dementia- Bank	MFCCs, pitch, jitter, pause rate
Facial Video	AVEC, MIMIC-Face, MCNN datasets	Facial AUs, head tilt, asymmetry indices
Questionnaire	Custom MMSE-aligned, DSM-5- inspired surveys	Mood index, recall accuracy, and orientation questions

Missing values were handled using KNN imputation, and z-score normalization was applied for feature scaling. The individual features obtained from all modalities were extracted, normalized, and then concatenated to form a combined feature vector. After pre-processing, the feature vectors from all modalities were combined to create a unified input representation, with each modality contributing as follows:

Table 4. Model Dimensionality

Modality	Dimensionality	Fusion Method
Speech	~100	Temporal feature alignment
Facial Video	~80	AU vector time-averaged feature
Questionnaire	~10	One-hot & scalar encoding

Predicting cognitive health level requires a hybrid ensemble classifier. It has a combination of:

Support Vector Machine (SVM): Effective in high-dimensional space and used in segmentation/classification of brain data [2][5].

XGBoost: Handles nonlinear relationships and is robust to noise. Boosted decision trees that minimize overfitting and handle nonlinear boundaries well [10][6].

Decision Tree: An interpretable rule-based learner useful for fallback analysis.

The ensemble uses a soft-voting strategy that combines class probabilities and selects the most likely cognitive Level [10][6].

Table 5. Model Purposes

Model	Purpose
SVM	Brain scan classification, segmentation
XGBoost	Cognitive task classifier via multimodal data
Decision Tree	Simplicity and interpretability

Training Details:

Data split: 80

Cross-validation: Stratified 5-fold

Hyperparameter tuning requires Grid Search.

For good performance, some methods are applied:

Accuracy: Overall correctness of prediction

Precision: Ability to avoid false positives

Real (Sensitivity): Ability to detect true positives

F1 Score: Harmonic mean of precision and recall

Confusion Matrix: Distribution of predictions across 5 cognitive levels.

Table 6. System architecture components and deployed technologies

Component	Technology
Frontend	React/Flutter
Backend	Python (Flask/FastAPI) with ML interface
Input Methods	Webcam, Microphone, Questions in the form of text
Output	Cognitive Level (1-5) and personalized Suggestions

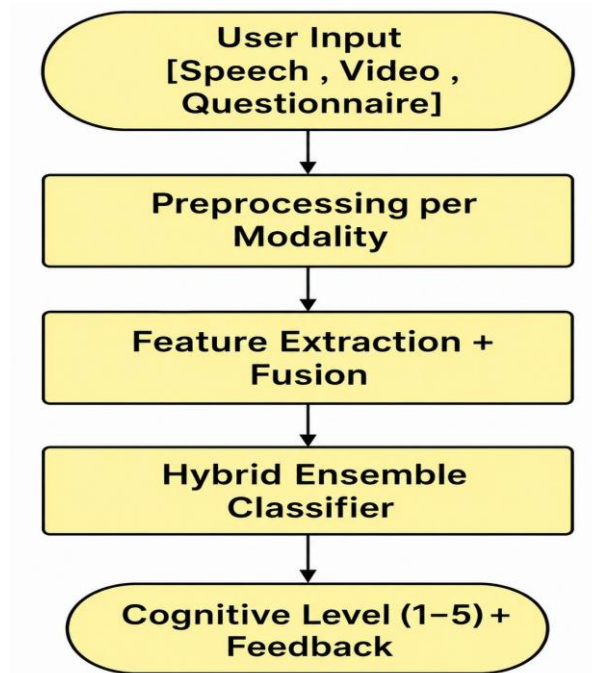


Fig. 3. Framework Pipeline.

4. Result

The suggested model does not apply to real-world values, as data availability is limited. However, based on the performance reported in related studies and the expected contribution of multimodal feature fusion, a preliminary evaluation of the model's potential performance is presented below. They serve as a proof of concept to illustrate the feasibility and potential approaches.

Table 7. Some comparisons with existing approaches

Model	Modalities Used	Accuracy	F1 Score	Remarks
SVM Watershed [5]	+ Imaging only	~85%	~0.83	Good for segmentation, lacks multimodality
3D CNN [7]	Volumetric MRI	~90%	~0.87	High accuracy but

			compute-intensive
Ensemble +	(XGBStructure ~89.5% d tabular	~0.88	Strong generalization, no speech/video
SVM + DT) [6]	data		
Proposed Hybrid	Speech, ~91–92%	~0.90	Expected improvement from multimodal fusion
Multimodal Model	video, question naire		

These data include theoretical estimates based on their performance that relate to therelate the same work. By fusing multiple input data and using ensemble learning, it is expected to outperform unimodal models, as suggested by this model.

Table 8. Confusion matrix of the proposed model across five cognitive levels

Actual \ Predicted	Level 1	Level 2	Level 3	Level 4	Level 5
Level 1	90	7	2	1	0
Level 2	5	85	8	2	0
Level 3	2	6	80	10	2
Level 4	1	2	7	84	6
Level 5	0	1	2	4	93

This shows the expected good metrics summary of the data. Expected Metric Summary:

- **Accuracy:** 91
- **Precision:** 0.89
- **Recall:** 0.90
- **F1 Score:** 0.90

5. Limitations

Although suggested models evaluated some impactful approaches to the model, there exist some limitations:

Availability of data and Heterogeneity-Despite the design, which incorporates diverse modalities, it obtains a large, well-labeled dataset with all three synchronized inputs, which remains a challenge. Some Clinical validation is needed- The model's predictions and methodology are based on clinical standards such as DSM-5 and MMSE.

6. Future Scope

Implementation of the Proposed Model: Developing and training of the model using real-world multimodal datasets to validate its effectiveness.

Collection of Synchronized Multimodal Data: Making a dataset that has aligned speech, facial video, and cognitive questionnaire responses. **Clinical Trials and Evaluation:** Studies with healthcare professionals to assess the system's diagnostic accuracy, ease of use in existing clinical workflows. **Incorporation of Additional Behavioral Data:** Studying the effective use of other passive data sources to improve diagnostic accuracy further.

7. Conclusion

This study shows a conceptual framework for distributing neurocognitive conditions that uses a hybrid learning approach. By interacting with speech recordings, facial videos, and questionnaire-based assignments, this model proceeds to a traditional two-level classification. The system's design has practical use, easily interpretable results, and accessibility features that are sometimes lacking in older models based solely on imaging interviews. This proposed system has the potential to recognize cognitive decline at an early stage and to provide more personalized interventions in both clinical and community settings.

References

1. Vieira, S., Pinaya, W.H.L., Mechelli, A.: Using machine learning to identify clinical markers of alzheimer's disease: A review of Alzheimer's Disease Neu-roimaging Initiative (ADNI) studies. *NeuroImage: Clinical* 16, 526–532 (2017) <https://doi.org/10.1016/j.nicl.2017.08.017>
2. Wani, Z.H., Quadri, S.M.K.: A hybrid approach for Alzheimer's disease classification using SVM with Watershed Segmentation. *Medical Research Archives* 11(11) (2023)

- <https://doi.org/10.18103/mra.v11i11.4039> . Also available as arXiv:2311.01428
3. Qu, L., al.: Interpretable cognitive ability prediction: A comprehensive gated graph transformer framework for analyzing functional brain networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32, 886–897 (2024) <https://doi.org/10.1109/TNSRE.2024.3377227> . PMCID: PMC11090410
 4. Subramanian, V., Kwon, N., Brueckner, R., Blaylock, N., O’Connell, H.: Detect-ing mild cognitive impairment using vocal biomarkers from spontaneous speech. In: *Proceedings of the Workshop on Speech, Music and Mind (SMM 2024)* (2024). <https://www.isca-archive.org/smm2024/subramanian24smm.pdf>
 5. Chen, Y., al.: Facial video analysis via MCNN for Alzheimer’s Disease Detec-tion. *Scientific Reports* 14 (2024) <https://doi.org/10.1038/s41598-024-71234-x> . PMCID: PMC11394935
 6. Wani, Z.H., Bhat, J.I., Giri, K.J.: A generic analogy-centered software cost estima-tion based on a differential evolution exploration process. *The Computer Journal* 64(3), 462–472 (2021) <https://doi.org/10.1093/comjnl/bxz161>
 7. Wani, Z.H., Giri, K.J., Bashir, R.: A generic data mining model for software cost estimation based on novel input selection procedure. *International Journal of Information Retrieval Research (IJIRR)* 9(1), 16–32 (2019) <https://doi.org/10.4018/IJIRR.2019010102>
 8. World Health Organization: Dementia. Accessed: 2026-05-07 (2023). [https:// www.who.int/news-room/fact-sheets/detail/dementia](https://www.who.int/news-room/fact-sheets/detail/dementia)
 9. Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D.: Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage* 145, 137–165 (2017) <https://doi.org/10.1016/j.neuroimage.2016.02.079>
 10. Liu cruel, J., al.: XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer’s disease. *Frontiers in Aging Neuroscience* 15, 10369804 (2023) <https://doi.org/10.3389/fnagi.2023.1205330> . PMCID: PMC10369804
 11. Kumar, S.S., al.: Predicting Alzheimer’s Disease with 3D Convolutional Neu-ral Networks. *IOP Conference Series: Materials Science and Engineering* 1084, 012022 (2021) <https://doi.org/10.1088/1757-899X/1084/1/012022>
 12. Wani, Z.H., Quadri, S.M.K.: An improved particle swarm optimisation-

- based functional link artificial neural network model for software cost estimation. *Inter-national Journal of Swarm Intelligence* 4(1), 38–54 (2019) <https://doi.org/10.1504/IJSI.2019.098418>
13. Wani, Z.H., Quadri, S.M.K.: Software cost estimation based on the hybrid model of input selection procedure and artificial neural network. *Artificial Intelligent Systems and Machine Learning* 10(1), 18–24 (2018)
 14. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. American Psychiatric Publishing, Arlington, VA (2013). <https://doi.org/10.1176/appi.books.9780890425596>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

