ISSN: 2093-5374 / E-ISSN: 2093-5382

A Deep Learning-based Psychometric Natural Language Processing for Credit Evaluation of Personal Characteristics

Dr.J. Praveenchandar[1*], S. Sankalp Karthi[2], R. Sowndharya[3], Dr.N. Dayanand Lal[4], Debarghya Biswas[5], and Manish Nandy[6]

[1*]Assistant Professor, Division of Artificial Intelligence and Machine Learning, School of Computer Science and Technology, Karunya Institute of Technology and Sciences (Deemed to be University), Coimbatore, India. praveenjpc@gmail.com, https://orcid.org/0000-0002-5735-8316

[2]Student, Department of Artificial Intelligence, Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, India. sankalp.karthi@gmail.com, https://orcid.org/0009-0007-9297-576X

[3]Assistant Professor, Department of Computer Science and Engineering, Sona College of Technology, Salem, India. rajendransowndharya@gmail.com, https://orcid.org/0009-0006-6209-0357

[4]Assistant Professor, Department of CSE, GITAM School of Technology, GITAM (Deemed to be University), Nagadenahalli, Doddabalapura, Bengaluru, India. dnarayan@gitam.edu, https://orcid.org/0000-0003-3485-9481

[5]Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India. ku.debarghyabiswas@kalingauniversity.ac.in, https://orcid.org/0009-0004-0730-9948

[6]Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India. ku.manishnandy@kalingauniversity.ac.in, https://orcid.org/0009-0003-7578-3505

Received: July 23, 2024; Revised: September 03, 2024; Accepted: October 02, 2024; Published: December 30, 2024

Abstract

Psychometric assessments that gauge individuals' skills, competencies, mindsets, and personality characteristics are essential for several practical applications, including online shopping, healthcare, and cybercrime. Conventional approaches cannot collect and measure extensive psychometric characteristics promptly and discreetly. As a result, despite their significance, psychometric factors have garnered less focus from the Natural Language Processing (NLP) and data sectors. This paper presents Deep Learning (DL), the Proposed model, designed to extract psychometric variables from user-generated texts. The proposed model incorporates an innovative representation insertion, a regional insertion, a Structural Equation Modelling (SEM) encoder, and a multitasking method, all functioning collaboratively to tackle the distinct issues of obtaining complex, nuanced, and user-focused psychometric measurements. The trials on three real-world datasets involving 11 psychometric characteristics, such as confidence, nervousness, and literacy, demonstrate that Proposed model significantly surpasses conventional feature-based classifications and leading DL frameworks. Ablation research indicates that every component of the Proposed model substantially

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 4 (December), pp. 151-165. DOI: *10.58346/JOWUA.2024.I4.010*

*Corresponding author: Assistant professor, Division of Artificial intelligence and Machine Learning, School of Computer Science and Technology, Karunya Institute of Technology and Sciences (Deemed to be University), Coimbatore, India.

enhances its general efficacy. The findings illustrate the effectiveness of the suggested design in enabling comprehensive psychometric examination.

Keywords: Psychometry, Natural Language Processing, Deep Learning, Credit Evaluation.

1 Introduction

Psychometrics pertains to the quantification of knowledge, abilities, attitudes, and personality features (Pellert et al., 2024). As the significance of micro-level analytics grows, particularly in comprehending and forecasting individual actions, the precise and prompt assessment of psychometrics is growing critically important. In information security, self-efficacy and danger perceptions are essential psychometric factors that significantly indicate end-user vulnerability to phishing attempts (Hooper & Blunt, 2020). Financial knowledge and psychological characteristics are significant precursors of subsequent financial actions (Lind et al., 2020). Contentment with a navigation capacity is a vital precursor for buying likelihood and electronic security (Yoo et al., 2023). In healthcare environments, psychometric assessments and feelings of security and fear concerning doctors significantly influence diverse health and wellness results, including future medical consultations and overall well-being (Netemeyer et al., 2020; Malathi et al., 2024). Thus, precise and fast measurement of psychometrics inside user-generated material dramatically enhances accessibility to data, benefiting various practical applications such as retrieving data, mobile text analysis, and modeling behaviors (Dover & Amichai-Hamburger, 2023).

Psychometric data collection has depended on monthly or quarterly survey-based methodologies. Successfully gathering and assessing pertinent information promptly and discreetly has been challenging in practical environments (Ashokka et al., 2020). Deep learning (DL) techniques for Natural Language Processing (NLP) are effectively utilized for psychometric aspects, including feelings and emotions (Gümüş et al., 2022; Kang et al., 2020). These NLP algorithms, which evaluate user-generated text and assign scores based on the target variable, enable real-time, passive monitoring and assessment.

Several problems hinder the extensive implementation of psychometric NLP methodologies for credit evaluation in environments (Walaa, 2024; Rathi et al., 2022). A limitation of psychometric NLP methodologies is that they can only reveal the author's personality through textual analysis, excluding insights into others, hence necessitating the preservation of the text's uniqueness. This assumption only sometimes applies in the tumultuous landscape of social media. Borrowers effortlessly replicate and paste information from other individuals to increase their credibility.

The research offers an innovative deep-learning framework for psychometric NLP to rectify these deficiencies (Ali & Raj, 2024; Ahmad et al., 2020). The design includes measures to tackle the above-described difficulties, using innovative representations and regional data, Structural Equation Modeling (SEM), and a multitasking approach. The suggested design was assessed using comprehensive healthcare testing equipment with different databases featuring relevant measures—healthcare literacy, math skills, trust, nervousness, and drug experiences—about a demographically varied customer group.

2 Background

Psychometric dimensions quantify latent components associated with knowledge, skills, views, opinions, and personality characteristics (Kyle et al., 2020). These aspects are recognized as significant antecedents, intermediaries, and modifiers for crucial humanistic results and actions. Health literacy is a subjective assessment of an individual's "knowledge regarding healthcare matters (Veera Boopathy et al., 2024)." Clients' confidence in physicians and fear of visiting them further illustrate health-related psychometric aspects. All three factors and other associated characteristics are demonstrated to influence healthcare results. The efficient collection and measurement of such variables in a timely and discreet way has remained challenging in practical contexts. Numerous psychometric variables need ten or more survey replies, rendering them less functional in continuous assessment contexts.

Recent research indicates that NLP techniques used with user-generated material provide a supplementary or alternative approach for assessing psychometric aspects (Kučera et al., 2020). Although NLP has a well-established history for particular characteristics like polarity of feelings (i.e., favorable, adverse, neutral) and specific feelings (e.g., satisfaction, rage), several significant psychometric aspects still need to be explained. This work aims to illustrate the effectiveness of NLP techniques in quantifying complex psychometric characteristics from text, considering the possible ramifications of psychometrics for data retrieval, as mentioned in the beginning.

Financial firms utilize in-house or outside credit scoring methods to assess their customer's credit risk, namely the likelihood of default or delinquency. Extensive efforts have been undertaken to determine people's credit risk over an extended period (Bhatore et al., 2020). This approach involves several models, namely two groups of methods: standard and sophisticated statistical models. Conventional statistical models encompass Naïve Bayes (NB), Logistic Regression (LR), and similar techniques. In contrast, advanced statistical modeling often pertains to information mining methodologies. Many elements for modeling are utilized. Credit risk evaluation systems typically include the present financial condition and past credit data (Zhang & Yu, 2024).

Significant work has been dedicated to enhancing various models using standard credit datasets to increase scoring capability. Sophisticated data mining models, such as Support Vector Machines (SVM), are employed (Kurani et al., 2023). Hybrid approaches have been developed to enhance scoring outcomes with increased accuracy. Zhang et al. introduced an innovative multi-stage hybrid credit scoring approach using a stacking-based ensemble technique (Zhang et al., 2021). Neural networks are typically employed in hybrid methodologies among these disciplines. Xu et al. developed a scoring methodology integrating case-oriented reasoning methodologies to mitigate Type I errors in the initial phase (Xu et al., 2023). SVM is utilized to create a hybrid approach. Sundararajan et al. utilized a varied weighted SVM and maximized the operating characteristic curve (Sundararajan & Srinivasan, 2023).

An effective credit evaluation model should incorporate borrowers' psychological tendencies and generate a holistic credit rating. DL techniques have been extensively utilized in the credit risk domain and have demonstrated satisfactory efficiency, yet the absence of explainability has constrained the advancement of these systems (Anitha Elavarasi & Jayanthi, 2022). The suggested methodology comprehensively leverages social media information to investigate its prediction capacity for personality characteristics.

3 Proposed DL-based Psychometric NLP Model for Credit Evaluation

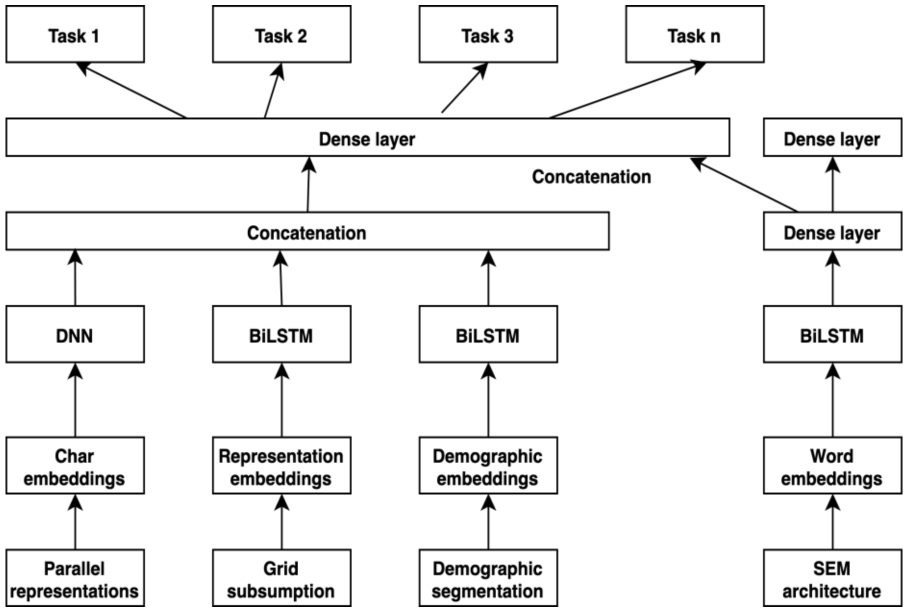

Figure 1: Workflow of the Proposed Model

Figure 1 illustrates the suggested Proposed model, Psychometric NLP DL Architectural Design, which integrates four foundational neural networks through a concatenation level that progresses into thick layers and employs an innovative multitask learning method. Each element of the design is designed to solve the previously listed deficiencies in research, hence improving text categorization skills for measurements:

Character Convolutional Neural Networks (CNN) is designed to capture essential spatial syntactic features produced by users' texts at the element, prefixing, suffixing, and rooting layers. A Bidirectional Long Short-Term Memory (Bi-LSTM) employs an innovative parallel representing encoding that integrates several subjects, sentiments, feelings, and grammatical linguistic expressions. This embedding utilizes characteristic subsumption techniques that assimilate extensive, varied characteristic fields and distill them into a concise collection of numerous characteristics.

- A secondary Bi-LSTM that integrates an innovative demographic embedding methodology designed to more effectively capture the subtleties and conventions present among various age, racial, and gender groups.
- An SEM Converter facilitates the incorporation of pertinent "secondary" behavioral and attitude data to enhance the categorization of essential target psychometric characteristics.
- A unique multitask learning technique that enhances the integration of joint data among related objective psychometric aspects.

3.1. Character Embedding

The research constructs a character-level embedding utilizing convolutional neural networks to analyze the input words' morphological features (e.g., prefixes, suffixes, and misspellings). Neural network-based insertions have demonstrated significant potential across several tasks. The input for a symbol embedding consists of a series of encoded letters. Every letter is denoted as a matrix $g(i) \in [1, l] \rightarrow R$, l represents the vocabulary dimension.

The character utilized in the model comprises 70 protagonists, including 26 English symbols, ten numerals, 33 additional symbols, and the fresh symbol. The kernel value is expressed as $f(i) \in [1, k] \rightarrow R$, k represents the filter size. With a stride of d, the combination of $h(j) \in [1, [1 - k + 1/l]] \rightarrow R$ among f(i) and g(i), which is obtained in Eq. (1).

$$h(j) = \sum_{i=0}^{k-1} f(i) g(j.d - i + c) \tag{1}$$

Where $c = k - l + 1$ is a variable. This structure is subsequently linked to a maximum pooling level, shown below in Eq. (2).

$$h(j)_{m-p} = \max\{g(j.d - i + c)\} \tag{2}$$

The embedding procedure employs two levels, every succeeded by a maximum pooling level. The resultant model is input into different ultimately linked levels combined with levels from additional insertions for the final categorization of psychometric variables.

3.2. Representation Embedding

Analyzing complex psychometric factors related to various consumer demographics provides difficulties for DL techniques, especially in contexts with restricted user-generated text. Recent studies indicate that sophisticated feature-based approaches frequently achieve text categorization performance similar to basic DL designs, while their integration often results in improved performance. The research offers innovative representation implantation that employs a diverse set of parallel characteristic depictions, capturing many semantics and syntactic data at different granularities and grid-based subordination. The primary concept underlying the suggested embedding resembles conventional word embeddings: to establish a lower-dimensional feature space that encapsulates essential features. As the research subsequently demonstrates scientifically, the representational embedding is exceptionally conducive to psychometric NLP, offering significant discriminating capability.

Grid-Based Subsumption (GBS)

While parallel models facilitate the incorporation of diverse language models at different granularities, they introduce the possibility of noise, redundant systems, and extraneous information. Some of the emphasized parallel data need more uniqueness or use. Previous research has suggested employing subsumption approaches to address this issue: character space reduction strategies mainly designed for natural language data. Previous approaches employ limited, predetermined subsumption procedures that need more scalability and extensibility in vast, dynamic characteristic fields. To address these constraints, the research offers an innovative GBS strategy adept at "reducing the grain from the chaff" utilizing the extensive parallel models.

Stage 1 of GBS mostly aligns with previous subsumption techniques, retaining only higher-order n-grams with superior discriminating capacity than characteristic equivalents inside the model. Considering the collection of m depictions R = $\{r_1, r_2, \cdots, r_n\}$,, when each r_k denotes a parallel illustration (e.g., word), the research derives all n-gram characteristics so that any f_{xyz} component of the characteristic set F corresponds to the xth characteristic in the n-gram group y for illustration r_z, and f_{xyz} first ranked in the following way in Eq. (3):

$$w(f_{xyz}) = \max\left\{p(f_{xyz}|c_a) \log\left(\frac{p(f_{xyz}|c_a)}{p(f_{xyz}|c_b)}\right)\right\} \tag{3}$$

Let c_a and c_b represent distinct elements from the set of C classification labels, where $c_a \neq c_b$. Let y denote the tokens in f_{xyz}, which possesses w potential character sensations. The s signifies the average angle value for every sense, calculated as the disparity among the favorable and adverse polarity ratings for reason q of f_{xyz} in SentiWordNet in Eq. (4).

$$s(f_{xyz}) = \sum_{j=0}^{d-1} \sum_{q=0}^{w-1} \frac{pos(f_{xyz},q) - neg(f_{xyz},q)}{dw} \tag{4}$$

The weighting equation's initial component evaluates the feature's discriminatory capacity through its log-likelihood ratio. In contrast, the subsequent component incorporates semantics to distinguish features with contrasting orientations (e.g., "like" versus "dislike") regarding their overall loads and subordination determinations. After weighing the features, the inside representation r is combined in Eq. (5). Every n-gram characteristic $f_{xyz}$with w(f_{xyz}) > 0 is juxtaposed to every lower-order n-gram characteristic f_{klz}, l < y, w(f_{klz}) > 0, and f_{klz}encompasses a subsection of tokens from f_{xyz}. If c(f_{xyz}) = c(f_{klz}), where:

$$c(f_{xyz}) = argmax \left\{ p(f_{xyz}|c_a) \log\left(\frac{p(f_{xyz}|c_a)}{p(f_{xyz}|c_b)}\right) + s(f_{xyz}) \right\} \tag{5}$$

The research then ascertains whether to include the higher-order n-gram based on the subordination limit, t shows in Eq. (6):

$$w(f_{xyz}) = \begin{cases} 1 & if \frac{H(G|r)}{H(G)} < d \\ 0 & else \end{cases} \tag{6}$$

H(G) denotes the throughout groups and $H(G \mid r)$ represents:

$$H(G|r) = -\sum_{r=0}^{N-1} P(r) \sum_{g=0}^{G-1} P(g|r) \log(P(g|r)) \tag{7}$$

In Stage 3, after establishing linkages between representatives as outlined in Equation (7), cross-representation subordination is executed for every pair of r_i and r_k where L(r_i, r_k) = 1. Given the bidirectional nature of the linkages, L(r_i, r_k)= L(r_k, r_i). Two-way contrasts are conducted, wherein each staying f_{xyz} with w(f_{xyz}) > 0 in r_i is juxtaposed to every characteristic f_{klz}, v < y, w(f_{klz}) > 0, f_{klz}encompasses a portion of tokens from f_{xyz}. Every remaining $f_{klz}$with w(f_{klz}) > 0 in r_k is analysed toward its equivalents in r_i that satisfy the conditions.

In phase four, the research considers strongly linked non-subsuming cross-validation characteristics. For every combination of r_i and r_k L(r_i, r_k)= 1, any surviving fix with w(f_{xyz}) > 0 in r_i is contrasted against every surviving f_{klz}in r with a weight larger than 0, where y = v. If the relationship among $f_{xyz}$and f_{klz}exceeds limit p, then w(f_{xyz}) = 0.

Embedding and BiLSTM

For every participation, the research employs word2vec to derive an I-sized embedding matrix for every token inside the corresponding dataset. Only tokens for which w(f_{klz}) = 0 are considered. A composite vector substitutes the embedding matrix for every other token. The embedding subsequently input into a Bi-LSTM to acquire the dependencies. In a series of phrases $w_1, w_2, \cdots, w_T$, which w_T represents a word embedded vector, the RNN acquires the latent properties of every word by considering all preceding words in the chain in Eq. (8).

$$h_p = \sigma(W_{hh} h_{p+1} + W_{hi} i_p + b) \tag{8}$$

Here, W_{hh} represents the weights matrix derived from the preceding concealed characteristics h_{p+1}, whereas W_{hh} denotes the weights vector associated with the source i_p,. Additionally, b signifies a biasing condition, and o indicates an operator. The formula used identifies latent properties derived from preceding words. The research generates concealed characteristics by deriving characteristics from subsequent words articulated as follows in Eq. (9):

$$h_p = \sigma\left(W_{hh}h_{p+1} + W_{hi}i_p + b\right) \tag{9}$$

RNNs capture long-term relationships; however, doing this in practice is challenging because of the gradient issue. LSTM employs inputs, forgets, and outputs gating to sustain longer enduring storage and storing dependency over time. It is structured as follows in Eq. (10) – (15):

$$x_p = \sigma\left(W(x)x_p + V(x)h_{p-1}\right) \tag{10}$$

$$f_p = \sigma\left(W(f)x_p + V(f)h_{p-1}\right) \tag{11}$$

$$o_p = \sigma\left(W(o)x_p + V(o)h_{p-1}\right) \tag{12}$$

$$\hat{c}_p = tanh\left(W(c)x_p + V(c)h_{p-1}\right) \tag{13}$$

$$c_p = f_p.c_{p-1} + x_p.\hat{c}_p \tag{14}$$

$$th_p = f_p.\tanh\left(c_p\right) \tag{15}$$

$W(x), W(f), W(o), W(c), V(x), V(f), V(o) and\ V(c)$ are weight matrix vectors contingent upon the input vocabulary vector and prior concealed characteristics. The Bi-LSTM is subsequently combined with the concealed attributes of additional embeddings, along with a softmax training on matrices, wherein the value of "1" is substituted with w(r_k) for every word.

3.3. Demographic Details

Demographics significantly influence people's linguistic patterns and psychological traits. The research develops innovative demographics word embedding to encapsulate the subtleties and conventions intrinsic to various demographic groupings. The population embedding determines sections with the highest entropy for a specific psychometric measurement, allowing for demonstrating within compared to across these demographics to reduce systematic error and improve categorization efficacy by aligning insertions more closely with consumer intrinsic semantic intent.

The initial goal is to find demographic factors that substantially influence the psychometric aspects of interest. The research employs a decision tree approach to achieve this objective. Consider a database $\{d_1, d_2, \cdots, d_N, K)$, where $D = d_1, d_2, \cdots, d_N$ that represents the collection of input demographic characteristics and $K = \{k_1, k_2, \cdots, k_N\}$ denotes the desired psychometric categories. The decision tree segments this database S into subgroups utilizing "nodes" based on the input variable d_m at specified dividing levels $u \in U(d_m)$.

U(d_m) denotes the whole set of potential values for the characteristic am. The objective is to develop tree branches that offer discriminating capacity for a specified target category k_n. This research sought to employs an entropy-based data acquisition measure for node choosing. The entropy H for the desired class C, which has potential values $\{k_1, k_2, \cdots, k_N\}$ and the likelihood mass product P(C), is expressed as shows in Eq. (16):

$$H(k) = -\sum_{x=0}^{N-1} P(k_x)\log\left(P(k_x)\right) \tag{16}$$

Data gain quantifies the decrease in entropy for the targeted classes when the database is further partitioned by a new input variable d_m. Discretization is employed to maintain characteristics that the computation of data gain. The data gain from the introduction of a characteristic $d_m$can be described as shows in Eq. (17):

$$CG(K, d_m) = H(K) - H(K|d_m = u_m) \tag{17}$$

H(K) is the entropy of the grouping token K, and the secondary data represents the anticipated entropy after partitioning the database utilizing feature d_m at u_m. The research constructs two varieties of decision trees for the demography integration. The initial category employs all demographic characteristics, referred to as "international tree" T_g. The subsequent form comprises a set of "regional trees" T_{xy}, each omitting one of the demographic factors. Like the random forest method, these regional trees utilize a random selection of input qualities to mitigate potential reliance on a limited number of dominating features. The research uses a binary tree architecture to ensure computational feasibility and implement a depth variable $l = \{1,2,\cdots,L\}$ in regulating the tree dimension. The population trees are constructed as shown below in Eq. (18) – (19):

$$yoT_g = \{d_m = u_m | d \in D, ht(T_g) = l\} \tag{18}$$

$$T_{xy} = \{d_m = u_m | d \in D, \{d_n\}, ht(T_{xy}) = l\} \tag{19}$$

Where ht() is the dimensional factor, the significant factors influencing the categories are chosen according to element scoring S shows in Eq. (20).

$$S(d_m = u_m) = \frac{NA(d_m=u_m)}{H(K|d_m=u_m)} + \frac{N(d_m=u_m)}{N(S)} \tag{20}$$

$d_0 = u_0 | T_g$ represents the root node requirement for the global tree. In contrast, $r_{xy}(S_1, S_2, \cdots S_{N-1})$ it denotes the top N - 1 node ratings for the nearby trees. The demographic integration utilizes this data in the following manner:

Let m_n denote one of the many components in M. For each m_n, the research finds a subset of the people in the training set who meet that criterion and generates a subcorpus consisting solely of text from these people. The study employs word2vec to derive an S-dimensional word embedding matrix for every word y in the subcorpus m_n, denoted as $w_{xy} = \{w_{xy1}, w_{xy2}, \cdots, w_{xyd}\}$.

For each person, the research delineates the subset $〚SS_D = \{ss_1, ss_2, \cdots, ss_d | ss \in SS\}$ of demographic data pertinent to that user. By the concept of average insertion, the socioeconomic embedding weight l_{xy} for phrase y, which occurs in a text occurrence linked to person v_x, is characterized as a weighted mean of the node rating S_w and the node-specific term integrating w_{sy} it shows in Eq. (21):

$$l_{xy} = \begin{cases} \frac{\sum_{x=0}^{SS_D-1} S_x w_{sy}}{|SS_D|} & if\ |SS_D| > 0 \\ l_{xy} & else \end{cases} \tag{21}$$

As previously said at the outset of this section, demographics embeddings enable the calibration of customer word embeddings according to the customer's statistics, illustrated with a tiny real-data example. The graphics depicted in Figure 3 are inherently symbolic, not comprehensive.

3.4. Structural Multitasking Operation

The user-based approach of psychometric research allows for linkages between targeted measurements, presenting a distinctive possibility for multitasking training. If "trust in physicians" and "depression of visiting doctors" are associated, the research combines their source's characteristics and concurrently trains the two classifications to enhance the group of attributes for the present task. The link between psychometric measurements has a tree-based model, as elaborated in the SEM model. The research constructs a representation that shares its structural characteristics to embody this distinctive attribute.

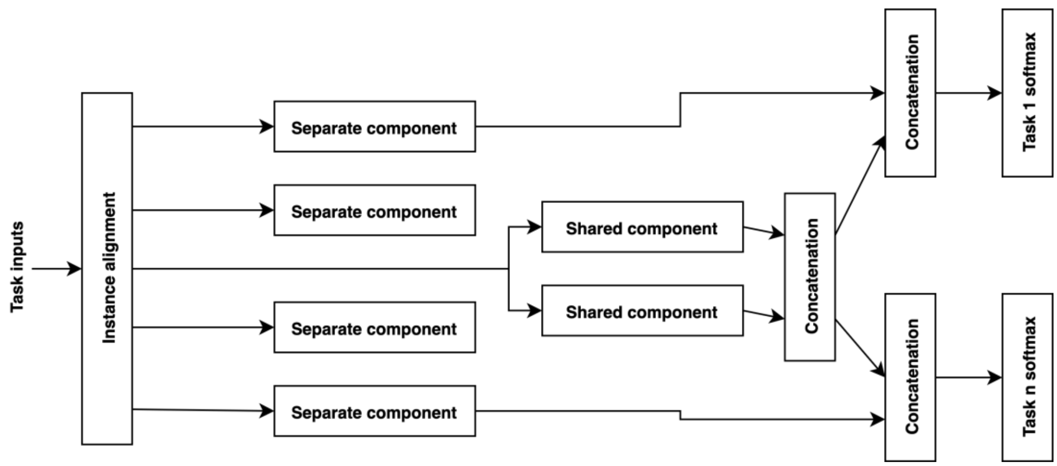

Figure 2: Structural Multitasking Model

Figure 2 illustrates the suggested multi-task learning strategy. Assume the research owns four desired variables of interest and intends to share their characteristics. The research developed several LSTMs, including task-specific characteristics alongside a singular "shared LSTM," which serves as cross-task participation, encapsulating similar patterns and signals across various categorization tasks. The research collaboratively trains these classifiers to facilitate feature exchange. Combining instruction maximizes the purity of sharing depictions, preserving orthogonality between sharing and independent models.

The objective is to construct two neural networks, a generator and a mechanism for discrimination, to compete against each other. The engine provides the most refined standard characteristic group, while the function seeks to differentiate the characteristics into distinct jobs. The solitary "common LSTM" functions as the power source, operating adversarially with a multilayer perceptron as the discrimination. Variations from the discrimination are required to enable the generator to refine the shared characteristic collection, which is transmitted back through a "Gradient Reversal Level (GRL)." GRL implements a matching product on the sources during the forwarding motion and transmits the adverse slopes in the reverse progress, facilitating training inside the generating system. To educate the discrimination, the research directly converts the shared models into a distribution of probabilities that forecasts the types of tasks that are inferred. The discriminator is trained via cross-entropy loss.

4 Results and Discussions

To evaluate the efficacy of the suggested design, the research performed a comprehensive benchmark assessment against 16 text categorization methodologies. The comparative techniques are categorized into five types: feature-based classification techniques (CNNs, Long Short-Term Memory Networks (LSTMs), hybrid DL architectures, and multitask DL approaches. The chosen techniques, albeit not comprehensive, exemplify cutting-edge methodologies within the five areas. Established feature-based classification algorithms, including Multinomial NB, LR, FastText, and Linear SVM, were added. To enhance the assessment, the research developed several bespoke CNN designs, including CNN-Word-Rep, which employs CNN to generate words and insertions after their integration into the thick levels for categorization, and CNN-Combine, which concurrently constructs word and illustration insertions using CNN. Notable LSTM designs were chosen, including the standard LSTM, LSTM-Word-Rep, and LSTM-Combine. The LSTMs-Then-CNNs approach was used as a mixed deep-learning infrastructure.

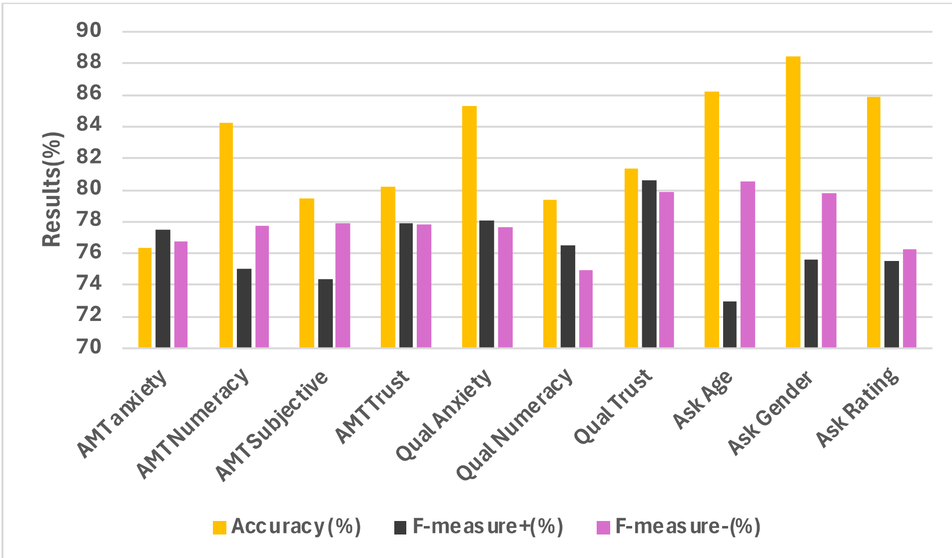

Figure 3: Psychometric Task Categorization Results

Figure 3 illustrates the accuracy and F-measures of the optimal technique for every group, categorized by the 11 psychometric task categories throughout the three datasets. Overall, the Proposed model surpassed the second-best approaches on 10 out of 11 correctness measurements, achieved superior performance on all F-measure metrics for the positive classroom, and excelled on 10 for the harmful category. The results together demonstrate the effectiveness and utility of the suggested design. The findings indicate that integrating CNN, LSTM, and multitasking with comprehensive insertions and encoding devices provides vital categorization accuracy across various datasets and psychological tasks such as classification.

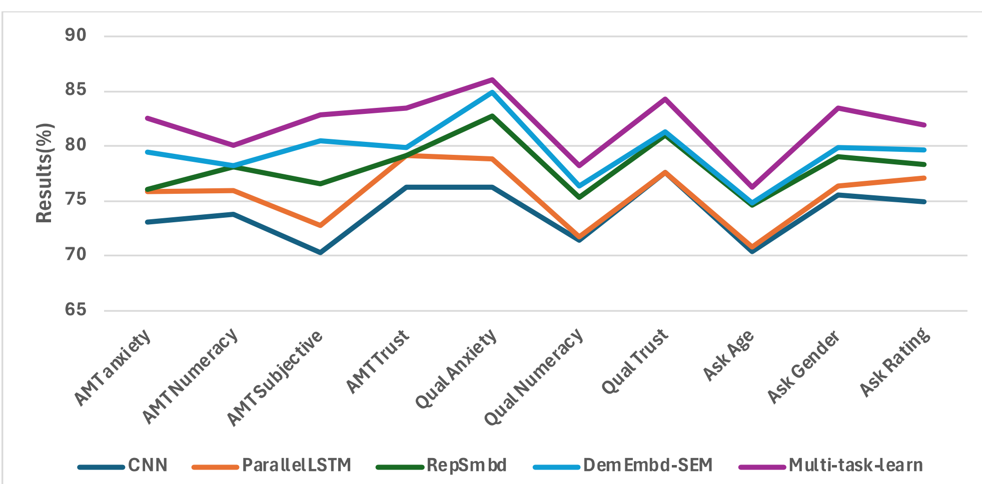

Figure 4: Ablation Assessment Analysis

Figure 4 illustrates the accuracies of ablation assessment for each of the 11 tasks across the trio of databases. The x-dimension demonstrates the effects of the five parameters. The research notes a consistent rising trend in nearly every job as supplementary elements of the Proposed model are progressively integrated. Although not illustrated here owing to spatial limitations, analogous plots were pointed out for the F-measures of both positive and negative classes. The picture demonstrates the resilience of every Proposed model element's cumulative involvement over the 11 activities.

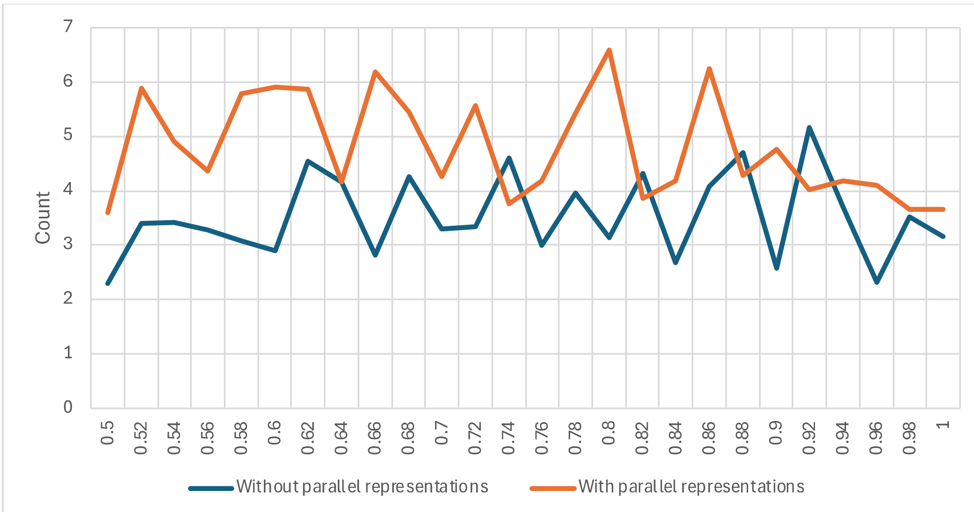

Figure 5: Graphical Representation Results

A crucial element of the proposed model is the abstractions and the associated representations anchoring that use GBS to enhance syntactical and aesthetic values in the area. Figure 5 illustrates using graphical anchoring to enhance Bi-LSTM ratings for genuine optimistic and damaging situations. The Bi-LSTM classification exhibits increased confidence in its precise forecasts, indicating that the parallel models successfully improve normalization.

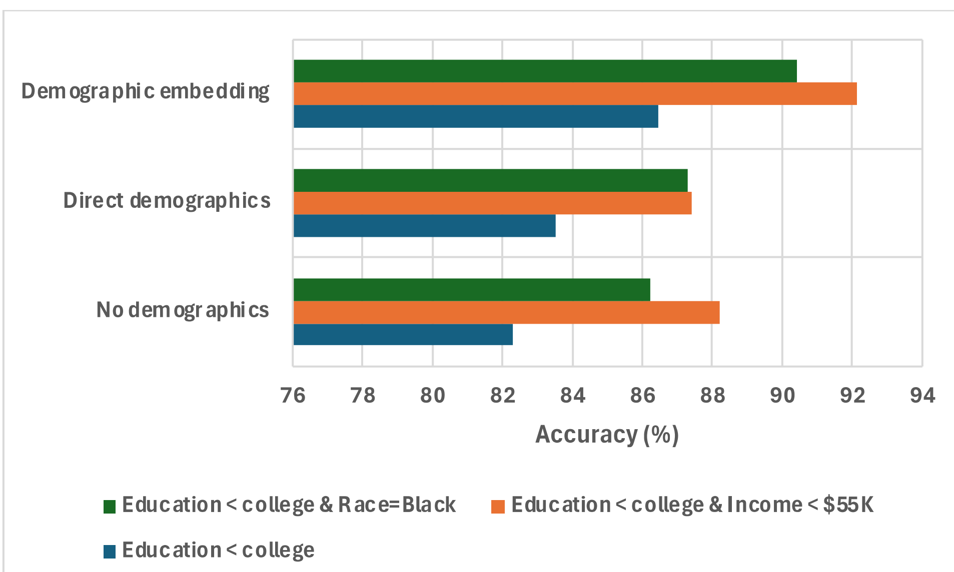

Figure 6: Disparity Analysis

As depicted in Figure 6, the disparities are particularly severe for several major statistical (and other) minority groupings within the information. The findings further emphasize the effectiveness of the splitting depiction utilized by the suggested model in improving prediction accuracy while mitigating bias. Replacing the "Parallel-Reps-LSTM" in the proposed model with previously trained embedding weights resulted in a decline in accuracy, precision, and recall by 1.5% to 2%, underscoring the significance of representative embeddings using LSTMs compared to previously trained embedded data.

The segment-specific insertions comprised 10% to 20% of the individuals in the sample. The demographic insertions for the numeracy test assigned varying weights to the words “capable,” “able,” “curiosity,” “comprehend,” “capacity,” and “complicated.” Likewise, terms like “anxious” and “concerned” were assigned varying weights when articulated by these segments within the categorization framework. The findings support the idea that adjusting the discriminating capacity of statements according to demographic factors might reduce bias and improve categorization precision. The ablation assessment findings emphasize the resilience of the proposed model’s embedded data, encoder, and multitasking learning framework.

5 Conclusion

Psychometric assessments that gauge individuals’ knowledge, abilities, attitudes, and personality characteristics have significant consequences for several critical real-world issues, including online shopping, medical care, and cybercrime. This article introduces a new DL construction, the Proposed model, to extract essential psychometric aspects like math, literacy, confidence, nervousness, and evaluations from NLPs. To mitigate the scarcity of inputs and to embody the awareness and user-based attributes of feature extraction, the proposed model comprises multiple suggested elements, including an illustration insertion, an insertion of SEM encoders, and a multitasking system. The trials across 11 tasks, including three datasets, demonstrate that the proposed model significantly surpasses conventional feature-based classifications and leading DL models.

The research enhances the existing literature on modeling users by including demographically calibrated embedded data and modeling with structural equations principles into text extraction and classification frameworks. The proposed model has significant practical ramifications; for instance, it dedicates users’ psychometric views and opinions that influence essential actions across numerous crucial domains, including health, cyber-security, and electronic commerce. In the healthcare sector, such models could be implemented through mobile applications to more promptly assess patients’ mental conditions associated with chronic illnesses using writing produced on mobile devices, thus aiding physicians in making educated choices and enabling patients to more effectively self-manage their health conditions. In future years, the research aspires to expand the concept to further specified application areas and implement it in real-time asynchronous chat environments. This work represents a significant preliminary advancement toward future practical applications.

References

[1] Ahmad, F., Abbasi, A., Li, J., Dobolyi, D. G., Netemeyer, R. G., Clifford, G. D., & Chen, H. (2020). A deep learning architecture for psychometric natural language processing. *ACM Transactions on Information Systems (TOIS)*, *38*(1), 1-29. https://doi.org/10.1145/3365211

[2] Ali, R., & Raj, M. (2024). Evaluating the Efficacy of Translanguaging Approach for Language Learning through K-Means Clustering Analysis. *Indian Journal of Information Sources and Services*, *14*(3), 232–240. https://doi.org/10.51983/ijiss-2024.14.3.30

[3] Anitha Elavarasi, S., & Jayanthi, J. (2022). Role of Machine Learning and Deep Learning in Assisting the Special Children’s Learning Process. *Journal of Algebraic Statistics*, *13*(2), 2327-2334.
[4] Ashokka, B., Dong, C., Law, L. S. C., Liaw, S. Y., Chen, F. G., & Samarasekera, D. D. (2020). A BEME systematic review of teaching interventions to equip medical students and residents in early recognition and prompt escalation of acute clinical deteriorations: BEME Guide No. 62. *Medical Teacher*, *42*(7), 724-737. https://doi.org/10.1080/0142159X.2020.1763286
[5] Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. *Journal of Banking and Financial Technology*, *4*(1), 111-138. https://doi.org/10.1007/s42786-020-00020-3
[6] Dover, Y., & Amichai-Hamburger, Y. (2023). Characteristics of online user-generated text predict the emotional intelligence of individuals. *Scientific Reports*, *13*(1), 6778. https://doi.org/10.1038/s41598-023-33907-4
[7] Gümüş, A. E., Uyulan, Ç., & Guleken, Z. (2022). Detection of EEG Patterns for Induced Fear Emotion State via EMOTIV EEG Testbench. *Natural and Engineering Sciences*, *7*(2), 148-168. https://doi.org/10.28978/nesciences.1159248
[8] Hooper, V., & Blunt, C. (2020). Factors influencing the information security behaviour of IT employees. *Behaviour & Information Technology*, *39*(8), 862-874. https://doi.org/10.1080/0144929X.2019.1623322
[9] Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139-172. https://doi.org/10.1080/23270012.2020.1756939
[10] Kučera, D., Haviger, J., & Havigerová, J. M. (2020). Personality and text: Quantitative psycholinguistic analysis of a stylistically differentiated Czech Text. *Psychological Studies*, *65*, 336-348. https://doi.org/10.1007/s12646-020-00553-z
[11] Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183-208. https://doi.org/10.1007/s40745-021-00344-x
[12] Kyle, G., Landon, A., Vaske, J., & Wallen, K. (2020). Tools for assessing the psychometric adequacy of latent variables in conservation research. *Conservation Biology*, *34*(6), 1353-1363. https://doi.org/10.1111/cobi.13625
[13] Lind, T., Ahmed, A., Skagerlund, K., Strömbäck, C., Västfjäll, D., & Tinghög, G. (2020). Competence, confidence, and gender: The role of objective and subjective financial knowledge in household finance. *Journal of Family and Economic Issues*, *41*, 626-638. https://doi.org/10.1007/s10834-020-09678-9
[14] Malathi, K., Shruthi, S. N., Madhumitha, N., Sreelakshmi, S., Sathya, U., & Sangeetha, P. M. (2024). Medical Data Integration and Interoperability through Remote Monitoring of Healthcare Devices. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(2), 60-72. https://doi.org/10.58346/JOWUA.2024.I2.005
[15] Netemeyer, R. G., Dobolyi, D. G., Abbasi, A., Clifford, G., & Taylor, H. (2020). Health literacy, health numeracy, and trust in doctor: effects on key patient health outcomes. *Journal of Consumer Affairs*, *54*(1), 3-42. https://doi.org/10.1111/joca.12267
[16] Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., & Strohmaier, M. (2024). Ai psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science*, *19*(5), 808-826. https://doi.org/10.1177/17456916231214460
[17] Rathi, S., Verma, J. P., Jain, R., Nayyar, A., & Thakur, N. (2022). Psychometric profiling of individuals using Twitter profiles: A psychological Natural Language Processing based approach. *Concurrency and Computation: Practice and Experience*, *34*(19), e7029. https://doi.org/10.1002/cpe.7029

[18] Sundararajan, K., & Srinivasan, K. (2023). Feature-Weighting-Based Prediction of Drought Occurrence via Two-Stage Particle Swarm Optimization. *Sustainability*, *15*(2), 929. https://doi.org/10.3390/su15020929
[19] Veera Boopathy, E., Peer Mohamed Appa, M. A. Y., Pragadeswaran, S., Karthick Raja, D., Gowtham, M., Kishore, R., Vimalraj, P., & Vissnuvardhan, K. (2024). A Data Driven Approach through IOMT based Patient Healthcare Monitoring System. *Archives for Technical Sciences, 2*(31), 9-15. https://doi.org/10.70102/afts.2024.1631.009
[20] Walaa, S. I. (2024). Threat Detection and Response Using AI and NLP in Cybersecurity. *Journal of Internet Services and Information Security*, *14*(1), 195-205. http://doi.org/10.58346/JISIS.2024.I1.013
[21] Xu, C., Chang, W., & Liu, W. (2023). Data-driven decision model based on local two-stage weighted ensemble learning. *Annals of Operations Research*, *325*(2), 995-1028. https://doi.org/10.1007/s10479-022-04599-2
[22] Yoo, S., Lee, D. J., & Atamja, L. (2023). Influence of online information quality and website design on user shopping loyalty in the context of e-commerce shopping malls in Korea. *Sustainability*, *15*(4), 3560. https://doi.org/10.3390/su15043560
[23] Zhang, W., Yang, D., Zhang, S., Ablanedo-Rosas, J. H., Wu, X., & Lou, Y. (2021). A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Systems with Applications*, *165*, 113872. https://doi.org/10.1016/j.eswa.2020.113872
[24] Zhang, X., & Yu, L. (2024). Consumer credit risk assessment: A review from the state-of-the-art classification algorithms, data traits, and learning methods. *Expert Systems with Applications*, *237*, 121484. https://doi.org/10.1016/j.eswa.2023.121484

Authors Biography

Dr.J. Praveenchandar, B.E., ME., PhD. is working as an Assistance Professor (selection Grade) in the Division of Artificial Intelligence and Machine Learning, under the School of Computer Science and Technology in Karunya Institute of Technology and Sciences (Deemed to be University), Coimbatore. He Completed his BE (CSE), ME (CSE) & PhD (ICE) in Anna University, Chennai. He got more than 17 years of teaching experience in various Leading Universities in India and his research area is Cloud computing and Machine Learning. He presented many of his research articles in International conferences. He published many research articles in SCI /Scopus/ Web of Science indexed journals. He is guiding many PhD research scholars in the field of Computer Science and Engineering, Artificial Intelligence, IOT etc. He own three International & Indian design patents and copyrights. He is the reviewer of many SCI indexed Journals of the reputed publishers like Springer, Elsevier, Tailor & Francis etc.

S. Sankalp Karthi, is a final year Artificial Intelligence Engineering Undergrad at Amrita School of Artificial Intelligence. He has worked on more than 15 projects over the past 3 years, in which IoT based Application for Realtime Heat Wave prediction with Weather Prediction and Analysis, and Realtime Tamil Fingerspelling Recognition is one of his best works. Currently he is working on FPGA Accelerated ViT and an LLM for Agriculture Assistant Artificial Intelligent system.

R. Sowndharya, working as Assistant Professor at Sona College of Technology, Salem. She completed M.E (CSE), B.Tech (IT).She has received IIT Best Award 2017 For Presented paper titled “Cardio vascular Disease Prediction Using Fussy Logic Expert System” in 2ndNational Conference on Recent Advances in Electronics held at ICFAI Tech School, Hyderabad. She has presented paper in national and international Conference. She has Published Paper in International journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Journal of Internet Services and Information Security (JISIS), IEEE Xplore. She has worked as Network Management Engineer Intern in Telecom Domain (BSNL). She has participated in numerous Faculty Development Program to get updated with new technologies. She has completed industrial training in field of IOT with AI. She has done various NPTEL courses.

Dr.N. Dayanand Lal, is currently working as an Assistant Professor in the Computer Science and Engineering Department at GITAM School of technology GITAM (Deemed to be University) Bengaluru. His research interest is in the areas such as Cloud Computing and Cyber Security. He has published more than 30 papers in National and International Peer Review Journals and Conferences and 1 Funded a Research Project Granted by (the DST) Department of Science and Technology on “Information and Technology assisted Silk Cocoon farming for the upliftment of the socio-economic status of Backward community rural mass” and has 2 International Patent Granted and 3 National Patents Published. He is presently supervising 5 Ph.D. Scholars, and has more than 15 years of teaching experience.

Debarghya Biswas, is an Assistant Professor in the Department of Computer Science & Information Technology at Kalinga University, Raipur, India. Research interests spanning Artificial Intelligence, Machine Learning, Cybersecurity, and Data Science. With a strong commitment to education, he teaches various courses including programming and software engineering, while also engaging in curriculum development and innovative teaching practices.

Manish Nandy, is an Assistant Professor in the Department of Computer Science & Information Technology at Kalinga University, Raipur, India. He holds advanced degrees in Computer Science and brings a wealth of knowledge in areas such as Software Development, Data Analytics, and Cybersecurity. With a passion for teaching, he engages students through innovative instructional methods and hands-on projects, fostering a dynamic learning environment.