Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Novel Hybrid Method for Pronominal Anaphora Resolution in Hindi Text

Shreya Agarwal[1], Dr Prajna Jha[2], Dr Tanveer J Siddiqui
[1]*Department of Electronics and Communication, University of Allahabad, India*

ARTICLEINFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Effective Aanaphora Resolution (AR) is essential for computational linguistics, as it underpins coherent text analysis, information processing pipeline, and the development of advanced language technologies. Pronominal Anaphora Resolution plays a crucial role in analyzing and understanding large text collections, enabling discourse understanding systems and enhancing the performance of related applications like text summarization, sentiment analysis, and machine translation. This paper proposes and evaluates a novel hybrid method for Hindi AR. The proposed method uses the rule-based method to resolve reflexive and locative pronouns, whereas it uses supervised classifiers to resolve demonstrative and relative pronouns. We investigate two machine learning and one deep learning classifiers- the Distributed Random Forest classifier, Stacked ensemble classifier, and Multi-Layer Perceptron. The performance evaluation is done on two standard datasets: the Hindi tourism dataset and the Hindi Dependency Treebank Data (HDTB). The stacked ensemble model outperforms all other models investigated in this paper on the Hindi Tourism dataset with an accuracy of 76.33%. The deep learning model performs better than stacked ensemble and random forest on the HDTB dataset with an overall accuracy of 75.96%. The proposed hybrid method outperforms most of the earlier reported work on Hindi AR. This research demonstrates the potential of DL and ML classifiers in developing an automatic entity linking system in Hindi text, which is necessary for the correct semantic interpretation of text..

Keywords: Anaphora resolution, Discourse Analysis, Multi-Layer Perceptron, Stacked ensemble classifier, Hybrid Model, Hindi text.

I. INTRODUCTION

Natural languages make use of references to link elements of text together. This linking can be forward or backward. The phenomenon of referring an entity that has been previously introduced in the discourse is termed as anaphoric reference. The process of identifying the entity being referred, called antecedent, by an anaphor is termed as Anaphora Resolution (AR) [1]. Consider the following excerpt from our dataset:

1. जोधपुर शहर, राजस्थान राज्य के पश्चिमी भाग के केन्द्र में स्थित, राज्य का दूसरा सबसे बडा शहर है और यह महलों, किलों और मंदिरों के लिए लोकप्रिय है।

   (Jodhpur city, located in the center of the western part of the state of Rajasthan, is the second largest city in the state and is popular for its palaces, forts, and temples.)

2. यह पर्यटन स्थल घूमने के लिए अच्छा है|

   yahparyatansthalghoomanekeliyeachchhahai|

   This tourist spot is good to visit.

3. शहर की अर्थव्यवस्था कई उद्योगों पर निर्भर करती है, जिनमें हस्तकला, टैक्सटाइल और कुछ धातुओं पर आधारित उद्योग हैं|

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

(The city's economy relies on several industries, including handicrafts, textiles, and some metal-based industries.)

4. मरुस्थल के हृदय में स्थित राजस्थान का यह शहर, राजस्थान का शानदार, प्रभावशाली और अपरिवर्तनीय मुकुट है |

(This city of Rajasthan, located in the heart of the desert, is Rajasthan's magnificent, impressive, and immutable crown.)

यह (it) in sentence 1,यह (this) in sentence 2, शहर (city) in sentence 3, यह (this) in sentence 4, all refer to the entity जोधपुर (Jodhpur). Human being can effortlessly resolve these references, however, replicating this ability in machines is a considerable challenge. AR significantly contributes to the effectiveness of other NLP systems, particularly in tasks such as discourse understanding, text summarization, text categorization, and sentiment analysis [1].

Significant state-of-the-art results are achieved in English AR [2]. However, research on AR involving Indian languages is limited. Demand for AR in non English speaking countries [3] has recently gained traction within the NLP research community [4]. The lack of benchmark dataset and linguistic tools pose major hindrance [5]. We take a small step in this direction by proposing and evaluating a novel hybrid algorithm for resolving anaphoric references in Hindi. We limit our scope to pronominal references only. The entity being referred to - antecedent - can be a noun or Noun Phrase (NP). The types of pronouns being studied are: demonstrative pronouns, reflexive pronouns, locative pronouns and relative pronouns. The proposed method combines the use of rule-base and classification models to resolve these for categories of pronouns. It uses manually crafted rules to identify antecedents of reflexive and locative pronouns and a pre-trained classifier to resolve references involving demonstrative and relative pronouns. We investigate Machine Learning (ML) and Deep Learning (DL) classifiers. The experimental investigations are done on datasets created using publically available Hindi tourism [6] and Hindi Dependency Tree bank Data (HDTB) [7]. Both the datasets have been used in recent publications involving Hindi AR. The classifiers being investigated are Random forest classifier, Stacked Ensemble classifier and Multi-Layer Feed-Forward Artificial Neural Network (ANN), also known as Multi-Layer Perceptron (MLP). It is trained with stochastic gradient descent (SGD) using back-propagation (BP) algorithm. While DL methods have shown promise for tabular data analysis [8], their application to Hindi AR remains largely unexplored. MLP is anticipated to work well on tabular data [9]. This work is an attempt to bridge this gap. The paper makes significant contributions in the following aspects:

1. The paper combines rule-based and machine, deep learning methods for Hindi AR. It investigates two machine learning classifiers - the Distributed Random forest classifier and stack ensemble classifier – for resolving anaphors. These classifiers with the considered feature set have not been investigated earlier for Hindi AR.
2. We also investigate a deep learning classifier, MLP, to resolve demonstrative and relative pronouns. MLP has not been investigated earlier for Hindi AR.
3. The performance of the proposed work is evaluated on two domains of the Hindi discourse and compared with existing state-of-the-art results. The proposed work performs better than recently reported work on HDTB dataset after 5-fold cross-validation.

The rest of the paper is structured as follows: the following section reviews the previous research on Hindi AR. The subsequent section provides details of the proposed hybrid method. Section 4 discusses the experiments and results. Finally, section 5 concludes the paper in Hindi AR.

II. LITERATURE REVIEW

Hindi is a low-resource language. Previous work in closely related low-resource languages, including Assamese [10] and Punjabi [11], has demonstrated the viability of developing effective AR solutions. These findings provided valuable insights for our exploration of Hindi AR. The existing work in Hindi AR can be categorized into four main categories, namely the Rule-Based approach, the ML-based approach, the DL-based approach and the Hybrid approaches. A challenge inherent in reviewing Hindi AR literature stems from the inconsistent publication frequency, characterized by substantial temporal gaps between prominent works, as detailed in the background

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

table. This discontinuity, particularly relevant to Hindi Anaphora Resolution, complicates the selection of solely recent references. To provide a comprehensive overview of the field's advancements, it is necessary to incorporate studies from a wider historical period, due the limited journal publications within the past five years.

Dutta et al. [12] used a modified Hobbs' algorithm for the resolution of reflexive and possessive pronouns in the Hindi discourse. Hobbs' algorithm operates on the principle that the antecedent of an anaphoric expression is typically the most recent referent in the discourse. The recent referent is expected to be compatible with the anaphor's syntactic role and the grammatical role in case of reflexive and possessive nouns, that make use of karakas and vibhakti information.

Agarwal et al. [13] resolved inter-sentential anaphoric references. Backus-Naur form (BNF) was used to check the grammatical structure of the sentence. Constraint-based criteria and metadata associated with potential antecedents were then used to identify the antecedent for an encountered anaphora. They achieved an accuracy of 96% in case of simple short stories and achieved an accuracy of 80% in case of compound sentences.

Prasad et al. [14] introduced a new S-list algorithm for Hindi text AR. They employed a novel ranking method for potential noun phrase antecedents, incorporating grammatical roles to prioritize candidates with greater likelihood of being the correct referent.

Mujadia et al. [15] proposed a dependency structure-based approach for AR in spoken Hindi dialogue was reported in. They utilized paninian grammar framework for the resolution of anaphors in dialogs. Their reported accuracy for user-user interaction dialog was 64%, while an accuracy of 59% was obtained for play story dialog corpora.

Uppalapu et al. [16] resolved first-person pronouns, second-person pronouns and third-person pronouns (TPP) in the Hindi discourse. They utilized karaka relations for salience ranking among a list of potential antecedent candidates for a given anaphora. The list of potential antecedent candidates was finalized by employing a modified S-list algorithm specific to the grammatical structure of the Hindi language. Compound sentences in the Hindi document were broken into simple sentences in which an anaphora was present. The authors achieved an overall accuracy of 77.25% on random Hindi text.

Singh et al. [17] resolved anaphoric references in different genres of the Hindi discourse using the gazetteer method. They established the significance of recency and animacy agreement between the potential antecedent and the anaphora. The authors achieved an accuracy of 64.8% in Hindi stories, 62.6% in Hindi news articles, and 82.6% in biographies.

The VASISTH system proposed by Sobha et al. [18] deals with one pronouns, distributives, reciprocals and gaps in Hindi and Malayalam, utilizing limited parsing information and exploring the morphological richness of the Indian languages. The author stated an accuracy of 82% on a dataset comprising of short and simple stories in Hindi.

Sikdar et al. [19] proposed a joint ensemble model for mention identification and correct antecedent resolution for the encountered pronominal anaphora in ICON-2011 dataset for Hindi, Bengali and Tamil languages. The authors employed a multi-objective differential evolution technique to optimize AR evaluation metrics and attain an optimal set of features capable of efficient AR in the above-mentioned languages. An MUC score of 33.27% and BCUB score of 66.37% in ICON 2011 Hindi dataset was reported by the authors.

An unsupervised approach for co-reference resolution (CR) in the Hindi HDTB dataset is reported by ramrakhiyani et al. [20]. The unsupervised Markov Logic Network (MLN) used probabilistic graph models to encode linguistic knowledge relevant for CR in the Hindi domain. The authors reported a promising accuracy of 55.04% on 10 HDTB [6] news articles.

Sikdar et al. [21] proposed a Condition Random field (CRF), ML-based approach for AR in ICON-2011 Dataset in Hindi and Bengali languages. They used the state-of-the -art framework proposed by Soon et al. [22]. The authors used the Beautiful Anaphora Resolution toolkit (BART) and the CRF classifier on a set of features to identify the potential antecedents. CRF was also used for the identification of correct antecedent-anaphora pairs. The authors achieved an MUC score of 66.70% in Hindi ICON 2011 dataset.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Dakwale et al.[23] proposed a hybrid approach for AR in the HDTB Data. They proposed a rule-based method for the resolution of demonstrative, reflexive, locative and relative pronominal anaphoric references. They used dependency structure information for salience determination of the potential antecedents. The references unresolved by the rule-based methodology were further resolved through a decision tree classifier using the framework proposed in [22]. The authors obtained a 70% accuracy on HDTB data.

Mahato et al. [24] formulated 21 rules for AR in the Hindi dialogue dataset apart from general filtering constraints. They achieved an accuracy of 52% in TPP.

Agarwal et al. [25] proposed a hybrid methodology for resolution of demonstrative, reflexive, and relative pronouns. Random forest and rule-based methodologies were adopted for the resolution of the aforementioned anaphoric references. The authors achieved an accuracy of 82% on Hindi tourism discourse.

Agarwal et al. [26] proposed LLM based hybrid method for AR in Hindi Text. They utilized stacked ensemble architecture and rules for resolution of distributive, relative, reflexive and demonstrative pronoun in Hindi tourism and HDTB dataset. The work utilized vanilla prompt engineering for identification of potential antecedents. An accuracy of 77.205% and 73.503% was reported for resolving such pronouns in Hindi tourism and HDTB dataset.

The seminal and path-breaking work in [27] on Coreference resolution, an umbrella term for AR, that dealt with end-to-end CR, inspired and further fueled work in AR utilizing DL techniques. Specific to AR systems in Indo-Aryan language like Hindi, the work in [28] and [29] have laid the foundation for DL-based Hindi AR, showing promising and significant results in developing a successful antecedent identification system at discourse level.

Lata et al.[28] proposed a bi-affine classifier usingGated Recurring Unit (GRU) and Convolutional Neural Network (CNN) for CR in the Hindi text achieving an F-score measure of 55.47%. Singh et al. [29] proposed GRU based approach for AR in Hindi code-mixed data, sourced from FIRE forum of AR shared task, reporting an F-score of 21%.

A detailed survey of AR, relative to the Hindi language can be found in [30]. Prajapati et al. [31] and Mishra et al. [32], have eloquently highlighted the different factors and existing approaches proposed for Hindi AR in multiple text domains.

Our literature survey reveals that majority of the existing work for Hindi AR has used rule-based, supervised ML based CRF classifier, decision trees or hybrid methodologies. Limited work using DL algorithms is reported in Hindi AR. There is no prior work, to our knowledge, that explores the potential of stacked ensemble classifiers with such an extensive feature set and MLP model for AR inferences. We have introduced features to reduce the bias in training sample by introducing a classification imbalance factor which is not explored in [26]. Our discourse consists of compound and complex sentences with multiple subjects in a single sentence, unlike simple short stories where entity resolution is more challenging. Additional features in the ML-based and DL-based module are introduced to capture the grammatical variation of the Hindi language by expanding the Soon et.al framework [22], assisting in developing an efficient AR system. This paper introduces a comprehensive feature set and evaluates the proposed algorithms on two diverse text genres, thereby enhancing model generalizability. The proposed DL model is computationally less complex than the existing DL- based frameworks.

III. METHODS

Figure 1 outlines the proposed hybrid method for Hindi AR. Potential antecedents are termed as ‘mentions’ [21] and the directly resolvable pronouns under consideration are termed as pronominal anaphora alias anaphora. A hybrid approach is established considered to be a robust approach for NLP tasks in the low-resource languages like Hindi [40]. The data is first pre-processed and then fed to alinguistic processing module. The linguistic processing module applies a sequence of operations. This includes part-of-speech tagging, dependency parsing and morphological feature extraction. The part-of-speech tag and noun chunking information is used to identify pronominal references and potential antecedents.

Anaphors resolved in this work are pronouns, which can belong to the following four categories: demonstrative, reflexive, relative and locative pronouns. A list of reflexive, locative, demonstrative and relative pronoun in Hindi is

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

maintained. List_1 contains reflexive and locative pronouns. List_2 contains relative and demonstrative pronouns. The presence and type of an anaphora, while processing a document from left to right, is identified. The potential antecedents in our case are recoverable concrete noun phrases. The search space for the Reflexive and Locative pronoun is limited to the sentence in which the pronoun occurred while the search space for relative and demonstrative pronoun comprises of the sentence in which the pronoun occurred and the previous three sentences. We use a rule-based approach to resolve reflexive and locative pronouns while the problem of resolving demonstrative and relative pronoun is modeled as a classification task. We investigate three different classifier for resolving demonstrative and relative pronoun. To train these classifiers, we annotate the dataset with a set of features using an annotation guideline discussed in table 1. The document is modeled as feature vector similar to [22, 25]. For each pair of an anaphora ’P’ and a potential antecedent ’N’ within the search a set of features are extracted from the annotated dataset. Table 1 elaborates the features used in this work. Gender agreement as a feature between the potential antecedent and anaphora is not considered in this study as pronouns in Hindi language are gender neutral. Adding a verb semantic analysis module for Hindi gender disambiguation would increase the computational complexity of the method. An extensive coverage of probable features for AR can be found in [36]. The ground truth is added as the last feature which is 'true' if the antecedent-anaphora pair is the correct paired and 'false' otherwise. The dataset is divided into train, test and validation set. The classifiers are trained on this labeled dataset. The trained classifiers are applied to test dataset and results are recorded.

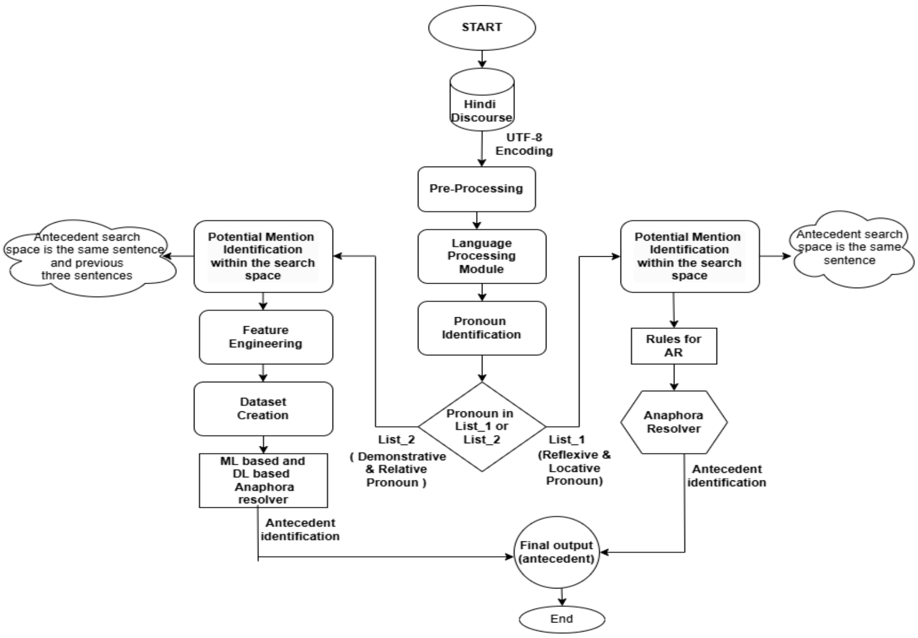

Fig 1: Functional Block diagram for AR in Hindi Text

A. *Data collection:* The proposed methodology is evaluated on two different dataset: a tourism dataset and HDTB dataset. In order to create tourism dataset, we extracted the raw corpus from publicly available Hindi Tourism dataset maintained by CFILT, IIT Bombay. This dataset contains annotations like, Part Of Speech tag, synset ids, etc. We converted it into plain text using python scripts and encoded it into UTF-8 format. The dataset was then annotated with useful features needed for Hindi AR task by adapting the annotation guidelines proposed in [34] and [35]. The corpus created from Hindi Tourism had 3.689k samples with 37 anaphora in list_1 and 349 anaphora from list_2. The corpus created from 48 documents from the HDTB on similar guidelines had 3.222k samples with 96 anaphora in list_1 and 344 anaphora from list_2.

B. *Preprocessing and Language Processing Module:* The data is first encoded into UTF-8 format and then pre-processed to remove the punctuation symbols, numbers, special symbols and other unwanted symbols. The pre-processed data is then passed to a linguistic processing module which uses stanza library [33] to perform the following operations:

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(i) Tokenize the text and assigns part-of-speech tag to each token

(ii) Identifies dependency relationship between words in a sentence using dependency parser. The resulting tree representation follows Universal Dependencies formalism performed by the 'DepparseProcessor' invoked with the name 'depparse'.

(iii) Extracts the morphological features of each word, like number, person using universal morphologicalfeatures (UFeats). This was performed along with the part-of- speech tagging by the POSProcessor and was invoked by name 'pos'.

This linguistic processing pipeline is not applied on the HDTB dataset as it already contains this information. The information extracted by the linguistic processing module is utilized in subsequent steps to identify potential antecedents.

C. *Pronoun Identification and Mention Identification Module:* The pronouns occurring in the dataset are easily identified using part-of-speech tags assigned by the POS tagger. We use a rule-based mention aaaaaaaaidentification module [37]. All the nouns and noun phrases appearing within the search space are considered as potential antecedent for an anaphora. The nouns and pronouns are identified using the part-of-speech tags while the noun phrases as potential antecedents are identified using the output ofthe dependency parser. Each word marked as "compound" by the dependency parser were clubbed together to form the multi-word entity. In Figure 2, गुलाबीनगर (Pink city), सिटीपैलेस (City Palace), महाराजाजयसिंहद्वितीय (King Jai Singh II), स्थापत्यकला (architecture) were easily identified as multiword noun by the dependency parser. Depending on the type of pronoun a rule-based module or machine and deep learning classifiers was invoked for resolution.

D. *Rule-Based AR module:* The reflexive and locative anaphors are resolved using manually crafted rules specific to the Hindi delineated in [18]. We demonstrate how the rules resolves reflexive and locative pronouns from our dataset.

Consider the following example(s):

1. "सवाई जय सिंह द्वितीय ने अपनी सिसोदिया रानी के लिये सिसोदिया रानी का बाघ बनवाया था |"

(Sawai Jai Singh II constructed the Sisodia Rani Bagh as a gift to his ownSisodia queen)

In the above sentence 'अपनी' (own)is a reflexive pronoun. 'जयसिंहद्वितीय'(Jai Singh II)and 'अपनी' (own)are singular and animate entities. 'सवाईजयसिंहद्वितीय' (Sawai Jai Singh II)is the subject of the clause in which 'अपनी' (own) reflexive pronoun occurs. Antecedent ' सवाईजयसिंहद्वितीय'(Sawai Jai Singh II is resolved as antecedent to the anaphora 'अपनी'(own).

2. "सवाई जय सिंह द्वितीय ने अपनी सिसोदिया रानी के लिये सिसोदिया रानी का बाघ बनवाया था |"

(Despite being landlocked, Bolivia is never suffocating, as here is stunning lakes and towering mountains)

In the above sentence 'यहां' (here) is a locative pronoun. 'बॉलिविया' (bolivia) precedes 'यहां'(here), 'बॉलिविया' (bolivia) is nearest to 'यहां'(here), 'बॉलिविया'' (bolivia) named entity category is place. Hence antecedent "बॉलिविया" (bolivia) is resolved as the antecedent to the anaphora 'यहां'(here).

E. *Machine Learning and Deep Learning Based AR Module:* The ML and DL classifiers are used for the resolution of relative and demonstrative pronouns. This is achieved by formulating the resolution task as a binary classification problem. Figure 2 shows the steps involved in ML and DL-based AR.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

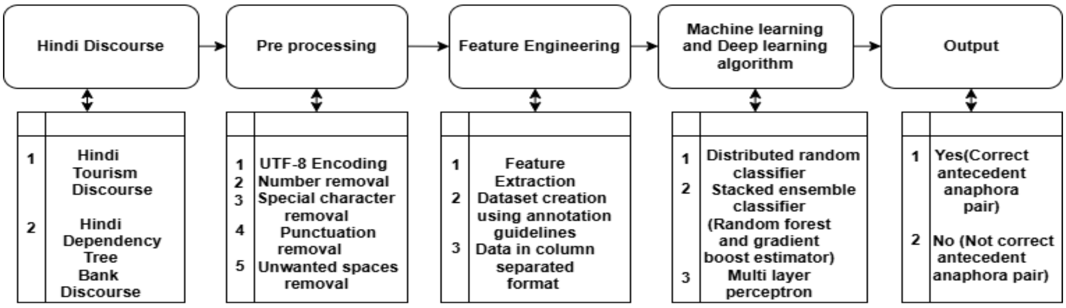

Fig 2: Flow of Classification process using Machine learning and deep learning algorithm.

Each instance in the training dataset is “<antecedent, anaphora>” pair. Each such pair is represented using a set of 13 features. These features contain numerical or categorical values of pre-defined attributes as discussed in Table 1. The classifiers are trained to maximize the marginal-likelihood of the correct antecedent-anaphora. The ML classifiers explored for AR are Random forest and the stacked ensemble model. The two classifiers explored for stacked ensemble model are Gradient Boost Estimator (GBE) and Random Forest Classifier (RF). Each layer receives the output from previous layer and computes weighted linear summation followed by some linear and non-linear processing. Given the increasing adoption of DL- based approaches in NLP applications [38], and inspired by the encouraging results achieved by DL-based AR methods for English [39], we further investigated a DL architecture [9] also for resolving demonstrative and relative pronominal references.The MLP framework had 3 hidden layers with 100 nodes, 10 nodes and 4 nodes respectively. The input was not raw Hindi data, but feature vectors and the output of the system was the final prediction whether the anaphora was paired with correct antecedent or not. The activation function used in all layers was the ‘tanh’ function with the gradient descent optimization technique and a drop out ratio of 0.1 to avoid over-fitting.

Table 1: Dataset features used in Hindi AR

S.no	Feature	Feature type	Feature value
1.	Number agreement	Categorical	Yes or No
2.	Person agreement	Categorical	Yes or No
3.	Animacy agreement	Categorical	Yes or No
4.	Sentence distance	Numerical	Integer
5.	Markable distance	Numerical	Integer
6.	Part of speech after the antecedent	Categorical	Part of Speech
7.	Part of Speech before the antecedent	Categorical	Part of Speech
8.	Grammatical role of antecedent	Categorical	Subject, object or indirect object
9.	Principal Noun or not	Categorical	Head or Modifier
10.	Type of anaphora	Categorical	Demonstrative

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

			or relative
11.	Direction of reference	Categorical	anaphoric or cataphoric
12.	Term frequency of antecedent	Numerical	Integer
13.	Class Distribution Aware Class Imbalance Factor	Numerical	Integer
14.	Target class	Categorical	Yes(pair corefer) or No (pair don't corefer)

Table 2: Results obtained on test data

Data-set	**Rule-based Accuracy**	**Classification model**	**AUC-PR**	**Overall accuracy**
Hindi tourism	89.16%	Random Forest	54.20%	**71.68%**
Hindi tourism	89.16%	Stacked Ensemble	63.50%	**76.33%**
Hindi tourism	89.16%	MLP	53.33%	**71.24%**
HDTB	84.33%	Stacked Ensemble	61.69%	**73.01%**
HDTB	84.33%	Random Forest	61.50%	**72.92%**
HDTB	84.33%	MLP	67.59%	**75.96%**

IV. RESULTS AND DISCUSSION

The proposed algorithms were evaluated on Tourism and HDTB dataset. Both the dataset is divided into 60:10:30 ratio for training, validation and testing purposes. Scripts in python language were utilized for the formulation of rules to resolve reflexive and locative pronouns. H20.ai framework was used for the development of the ML and DL models respectively. We conducted test runs on both the dataset. All the test runs were conducted on a machine equipped with a Core-I7 processor and 16 GB RAM. The results are taken using all the three classifiers on both the dataset. The performance of the proposed methods is assessed in terms of Precison, reccall, accuracy and Area under the Precision Recall Curve (AUC-PR) metrics.

Table 2 shows the results. From Table 3, it is evident that the stacked ensemble model exhibits better performance on Hindi tourism dataset while the MLP model outperforms all other models investigated in this work on HDTB dataset. The MLP model achieves an overall accuracy of 71.24% and 75.96% on the Hindi tourism and the HTDB dataset respectively. The overall accuracy observed using stacked ensemble model is 76.33% and 73.01% respectively.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3 compares the results obtained by the proposed method with the state-of-the art works in Hindi. As shown in the table, the proposed method performs better than most of the notable results reported for AR on the Hindi text. The ML-based stacked ensemble hybrid method yields the best results in correctly identifying anaphora in the Hindi tourism corpus with comparable performance by hybrid MLP-based approach on the Hindi tourism corpus. Exploring DL models on tabular data in Hindi AR is first of its kind attempt, and is believed to be a novel contribution of this paper.For the publicly available HDTB dataset, DL-based hybrid approach surpassed all existing work reported in AR on the HDTB corpus by achieving an accuracy of 75.96% resulting in 8.5% improvement over work reported in [23].

Table 3: Comparative analysis with existing work in Hindi AR

Author name[reference no]	**Methodology**	**Dataset**	**Evaluation results**
Uppaapu*et al.*[16]	Rule-based Modified S list algorithm	Long Hindi stories	64% accuracy
Sobha*et al.* [18]	Rule-based	3000 Hindi words	82% accuracy
Ramrakhiyani*et al.* [20]	Unsupervised Markov Logic Network	HDTB	55.04% accuracy
Sikdar*et al.* [21]	ML Conditional Random Field Classifier	ICON 2011	62.80% F-score
Dakwale*et al.* [23]	Hybrid(rules+ decision tree)	HDTB	70% accuracy
Mahato*et al.* [24]	ML classifier	1059 Hindi sentences	52% F-score
Agarwal *et al.* [25]	Hybrid (rules +random forest)	100 Hindi tourism sentences	82% accuracy
Lata*et al.* [28]	DL-based Bi-GRU-CNN	HDTB	55.47% F-score
Agarwal et al [26]	Hybrid (rules + stack ensemble)	HDTB Tourism	73.503% accuracy 77.205% accuracy
Proposed Methodology	Hybrid(rules +Random forest)	HDTB Tourism	72.92% accuracy 71.68% accuracy
Proposed Methodology	Hybrid(rules+ stacked ensemble)	HDTB Tourism	73.01% accuracy **76.33% accuracy**
Proposed Methodology	Hybrid (rules+ Multi-layer Perceptron)	HDTB Tourism	**75.96% accuracy** 71.24% accuracy

Agarwal et al. [26] did not dealt with locative pronouns in both genre of text, which we have considered in our study. We also found the findings of Singh et al. [41] true in Hindi tourism data as well as in the HDTB dataset. Authors in [41] argued that in a Hindi discourse, majority of the pronouns were demonstrative and were generally directly recoverable. Our model demonstrates certain limitations in efficiently resolving anaphoric references where antecedents are verb phrase or events [42] or the antecedent are cataphoric.

V. CONCLUSION

Despite significant advancements in the field of NLP, the task of AR is far from fully solved. AR is important for correct semantic interpretation of the text. This paper takes a step forward in this direction by developing an automatic AR systems for low-resource language like Hindi. Two different hybrid frameworks are experimented with for AR in two genres of the Hindi text. ML-based hybrid models used RF classifier and manually crafted rules. Another ML-based hybrid model experimented was stacked ensemble architecture and rules, achieving an accuracy of 76.33% on Hindi tourism corpus and 73.01% on the HDTB dataset. Deep tabular learning framework appended with rules for AR in the HDTB achieved significant accuracy resulting in the significant improvement over the existing methods. Our experimental results demonstrate a marked superiority in AR accuracy for Hindi news data, with the RF& rules, stack ensemble & rules, and deep tabular & rules models outperforming the state-of-the-art results in [23] by 4.1, 4.3, and 8.5 percentage points, respectively. The proposed approach efficiently deals with inter and intra-sentential pronominal anaphoric references.

REFERENCE

[1] Mitkov, R., Evans, R., Orăsan, C., Ha, L. A., &Pekar, V. (2007). Anaphora resolution: To what extent does it help NLP applications?. In Anaphora: Analysis, Algorithms and Applications: 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos, Portugal, March 29-30, 2007. Selected Papers 6 (pp. 179-190). Springer Berlin Heidelberg.Sukthanker, R., Poria, S., Cambria, E., Thirunavukarasu, R. Anaphora and coreference resolution: A review. Information Fusion. 2020, 59,139-62.

[2] Sukthanker, R., Poria, S., Cambria, E., &Thirunavukarasu, R. (2020). Anaphora and coreference resolution: A review. Information Fusion, 59, 139-162.

[3] Dongare, P. (2024, May). Creating Corpus of Low Resource Indian Languages for Natural Language Processing: Challenges and Opportunities. In Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation (pp. 54-58).

[4] Poesio, M., Ogrodniczuk, M., Ng, V., Pradhan, S., Yu, J., Moosavi, N. S., &Zeman, D. (2024, May). Universal anaphora: The first three years. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 17087-17100).

[5] Chhabra, A., Arora, M., Sharma, A., Singh, H., Verma, S., Jain, R., ...& Kanavos, A. (2024). Classifying hindi news using various machine learning and deep learning techniques. International Journal on Artificial Intelligence Tools, 33(02), 2350064.

[6] Bhingardive, S., Shukla, R., Saraswati, J., Kashyap, L., Singh, D., & Bhattacharyya, P. (2016, May). Synset Ranking of Hindi WordNet. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 3039-3043).

[7] Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., & Xia, F. (2009, August). A multi-representational and multi-layered treebank for hindi/urdu. In proceedings of the third linguistic annotation workshop (LAW III) (pp. 186-189).

[8] Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., &Kasneci, G. (2022). Deep neural networks and tabular data: A survey. IEEE transactions on neural networks and learning systems.

[9] Gorishniy, Y., Rubachev, I., Khrulkov, V., &Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in neural information processing systems, 34, 18932-18943.

[10] Das, M., &Senapati, A. (2025). Peer-to-Peer Network for Co-reference Resolution. Journal of Scientometric Research, 14(1), 342-350.

[11] Kaur, K., Goyal, V., & Dutta, K. (2021, December). Machine Learning approach for resolving anaphors in the Punjabi Language. In 2021 1st International Conference in Information and Computing Research (iCORE) (pp. 50-55). IEEE.

[12] Dutta, K., Prakash, N., & Kaushik, S. (2008). Resolving pronominal anaphora in hindi using hobbs algorithm. Web Journal of Formal Computation and Cognitive Linguistics, 1(10), 5607-11.

[13] Agarwal, S., Srivastava, M., Agarwal, P., &Sanyal, R. (2007, August). Anaphora resolution in hindi documents. In 2007 International Conference on Natural Language Processing and Knowledge Engineering (pp. 452-458). IEEE.
[14] Prasad, R., Strube, M. (2000). Discourse salience and pronoun resolution in Hindi, Current Work in Linguistics, University of Pennsylvania Working Papers in Linguistics, 2000.
[15] Mujadia, V., Agarwal, D., Mamidi, R., & Sharma, D. M. (2015, October). Paninian grammar based hindi dialogue anaphora resolution. In 2015 International Conference on Asian Language Processing (IALP) (pp. 53-56). IEEE.
[16] Uppalapu, B., & Sharma, D. M. (2009, April). Pronoun resolution for hindi. In 7th Discourse anaphora and anaphor resolution colloquium (DAARC 2009) (pp. 123-134).
[17] Singh, S., Lakhmani, P., Mathur, P., &Morwal, S. (2014). Anaphora resolution in hindi language using gazetteer method. International Journal on Computational Sciences and Applications IJCSA, 4, 567-9.
[18] Sobha,L., andPatnaik,B. Vasisth: An anaphora resolution system for Malayalam and Hindi (2000). In: Proceedings of Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications, Monastir, Tunisia..
[19] Sikdar, U. K., Ekbal, A., &Saha, S. (2016). A generalized framework for anaphora resolution in Indian languages. Knowledge-Based Systems, 109, 147-159.
[20] Ramrakhiyani, N., Hingmire, S., Pawar, S., Patil, S., Palshikar, G., Bhattacharyya, P., &Verma, V. (2018, December). Resolving actor coreferences in hindi narrative text. In Proceedings of the 15th International Conference on Natural Language Processing (pp. 50-58).
[21] Sikdar, U. K., Ekbal, A., Saha, S., Uryupina, O., &Poesio, M. (2013, October). Adapting a state-of-the-art anaphora resolution system for resource-poor language. In Proceedings of the sixth international joint conference on natural language processing (pp. 815-821).
[22] Soon, W. M., Ng, H. T., & Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. Computational linguistics, 27(4), 521-544.
[23] Dakwale, P., Mujadia, V., & Sharma, D. M. (2013, October). A hybrid approach for anaphora resolution in hindi. In Proceedings of the sixth international joint conference on natural language processing (pp. 977-981).
[24] Mahato, S., Thomas, A., &Sahu, N. (2020). Heuristic algorithm for resolving pronominal anaphora in Hindi dialects. In Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2018, Volume 1 (pp. 41-51). Springer Singapore.
[25] Agarwal, S., Jha, P., Abbas, A., & Siddiqui, T. J. (2023, March). Information Extraction for Design of a Multi-feature Hybrid Approach for Pronominal Anaphora Resolution in a Low Resource Language. In International Conference on Advanced Computing, Machine Learning, Robotics and Internet Technologies (pp. 169-180). Cham: Springer Nature Switzerland.
[26] Agarwal, S., Jha, P., Abbas, A., & Siddiqui, T. J. (2025). A Hybrid Stacked Ensemble Model for Resolving Pronominal Anaphoric Ambiguity in Hindi Discourse. SN Computer Science, 6(3), 248.
[27] Lee, K., He, L., Lewis, M.,Zettlemoyer, L. (2017,). End-to-end neural coreference resolution. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, pp. 188-197 .
[28] Lata,k., Singh,P., Dutta,K.,Kanwar,A. (2024). DeepHCoref: A Deep Neural Coreference Resolution for Hindi Text. In Proceedings of the Seventh Workshop on Computational Models of Reference, Anaphora and Coreference, Miami, Florida , pp. 67-77.
[29] Singh, S., Patel, K., Bhattacharyya, P. Attention-based Anaphora Resolution for Code-Mixed Social Media Text for Hindi Language. In FIRE (Working Notes), 2020, pp. 780-787.
[30] Mahato, S., Thomas, A., &Sahu, N. (2017). A relative study of factors and approaches for Hindi Anaphora resolution. International Journal of Management IT and Engineering, 7(12), 176-188.
[31] Prajapati, P., Goyal, V., & Kaur, K. (2024). A Detailed Study on Anaphora Resolution System for Asian Languages. SN Computer Science, 5(7), 811.
[32] Mishra R, Desur P, Shah RR, Kumaraguru P. (2024). Multilingual Coreference Resolution in Low-resource South Asian Languages. In Ptoceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC COLING), Torino, Italia , pp. 11813-11826.

[33] Qi, P., Zhang, Y., Zhang, Y., Bolton, J., Manning, C.D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. InProceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 101-108.
[34] Dakwale, P., Sharma, H, Himanshu, D.M. (2012). Anaphora annotation in Hindi dependency treebank. InProceedings of the 26th Pacific Asia Conference on Language, Information and Computation, Bali, Indonesia, pp. 391-400.
[35] Mujadia, V., Gupta, P., Sharma, D.M. (2016). Coreference annotation scheme and relation types for Hindi. InProceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portoroz, Slovenia, pp. 161-168.
[36] Lata, K., Singh, P., Dutta, K. (2021). A comprehensive review on feature set used for anaphora resolution. Artificial Intelligence Review, 54, 2917-3006.
[37] Lata, K., Singh, P., Dutta, K. (2023). Semi-automatic Annotation for Mentions in Hindi Text. SN Computer Science, 4(5), 515.
[38] Otter, D.W., Medina, J.R., Kalita, J.K. (2020). A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, 32(2), 604-24.
[39] Stylianou, N., Vlahavas, I. (2021). A neural entity coreference resolution review. Expert Systems with Applications, 168,114466.
[40] Jadon, A.S., Parmar, M., Agrawal, R. (2024). Hinglish Sentiment Analysis: Deep Learning Models for Nuanced Sentiment Classification in Multilingual Digital Communication. In 2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT), Dehradun, India, pp. 318-323.
[41] Singh, P., Dutta, K. (2015). Sensitivity analysis of feature set employed for anaphora resolution. International Journal of Computer Applications, 128 (14), 10-4.
[42] Mujadia, V., Gupta, P., Sharma, D.M. (2016). Pronominal Reference Type Identification and Event Anaphora Resolution for Hindi. Int. J. Comput. Linguistics Appl., 7(2), 45-63.