Bio-NER: Biomedical Named Entity Recognition using Rule-Based and Statistical Learners

Pir Dino Soomro, Sanotsh Kumar, Banbhrani, Arsalan Ali Shaikh, Hans Raj
School of Computer Science and Technology
Dalian University of Technology
Dalian, 116024, P. R. China

***Abstract*—The purpose of extracting of Bio-Medical Entities is to recognize the particular entities, whether word or phrases, from the unstructured data contained in the text. This work proposes different approaches and methods, i.e. Machine Learning Hybrid Classification, Rule Based Non-tested Generalized Exemplars and Partial Decision Tree (PART) Learners for Bio-Medical Named Entity Recognition. The Prime objective is to consider, preferably, simple characteristics, such as, affixes and context. In addition, orthographic, Parts of Speech (POS) tags and N-grams are given secondary importance as for as their comparison with affixes and context is concerned. Further, for the very purpose of Bio-medical Diseased Named Recognition, proposal of Rule Based Classifiers along with the Statistical Machine Learning is given. Also, this paper proposes the blend of both preceding methods that jointly construct Hybrid Classification algorithm. Precision, Recall and F-measure – standard metrics- has been put into practice for the evaluation. The results prove that the technique used has far better performance results than the method used before - state-of-art Disease NER (Named Entity Recognition).**

Keywords*—Bio-medical text mining; machine learning; named entity recognition; naive bayesian; rule-based classifier; information extraction

I. Introduction

Nowadays, in context of bio-medical domain, the bio medicinal work is going to increase rapidly because of the time, the developing measure of the content on World Wide Web (WWW). Internet, a viable and productive information recovery system, is required. So in bio-medical domain the bio medicinal work has been expanded; the measure of content in online sources i.e. MEDLINE, which is, as of now, the biggest archive for bio medical works. In biomedical work, namely, elements signifies to word or grouping of the word which represent particular terms, such as; protein, DNA, RNA or ailment name. Because of the enormous development of content, effective information recovery and automation is required. The procedure of labeling individual substances is called Named Entity Recognition (NER). And the NER is the most vital development in the extraction of learning, which has the general point of distinguishing particular terms like, Protein, Gene, Disease and medication [2]. Until in recent past, much consideration has been centered around NER of protein and gene items, while little work has been led on sickness NER [3]. Bio-NER has been difficult when contrasted with normal NER (Area, Names, Time, Date and so on). Execution of the (Bio-NER) contrasted with Named Entity Recognition, the Biomedical Named Entity Recognition is high because of the accompanying reasons [3], [6]. First, the elements of biomedical filed unavailability of a tenacious morphology and consequently, they are not a formal noun (people), places or things comprising letters, numbers and so on which, additionally, expanding disambiguate of grouping. Second, highest critical arrangement problem is the united conveyance of the content, for instance; Cancer can be delegated a modifier; it can be additionally named a particular ailment and malady class and so on.

Thus, we prevalently concentrate on Disease Name Recognition by utilizing the National Center for Biotechnology Information (NCBI) dataset in this examination. For this very reason, Rule Based Learners - (PART, DTNB and Non-Nested GE) - and Machine Learning Technique, for example, (Naive Bayesian, Bayesian Network) has been well-thought-out for Named Entity Recognition (NER). Performances of these classifiers were analyzed utilizing standard measurements such as; exactness accuracy, recall and F-score. Besides, the examination has been done to assess the combination of machine learning approach and rule based learners for Disease Name Recognition. The best in class, Statistical Machine Learning technique which demonstrated better performance above distinct statistical Machine Learning strategies, in the direction of perceiving illness named elements of biomedical work. Significantly, the prime focus is on Rule based methods (Partial Decision Tress, Naive Bayesian Decision Table and Non-Nested GE) and statistical learning (NB, BN) speculatively appropriate toward different Named Entity Recognition issues.

The rest of the paper discusses and runs ahead as: Section II presents NER utilizing Rule Based learners and Machine Learning; Section III represents the proposed technique, assortment of characteristic for ailment NER, and selection of the methods to do experimentation/test, we at that point present new technique classifier fusion method/technique. Section IV gives the visions about the test setup and talks about how Rule base learner and machine learning consolidated as well the data Sets utilized as a part of the investigation.

II. Related Work

Now, in existing time, there exist a massive amount of material, information and data existing in the form of Natural Language. IE is a range of research that conveys the design approach and usage of frameworks that assist automatically to remove specific sorts of organized data or material from

archives. Named Entity Recognition (NER) whilst it is a unit of (IE), the procedure Entity Extraction (EE) as well famous with (EE), which recognizes nuclear components in content and arranges or orders by classifying those components in the classifications which are established in advance [5]. Named entities is to mention name of individuals, association, area, position and so on, as opposed to common entities recognition Not much work is available in the biomedical field as for as this area is concerned. The removal of substances related with the bio-medical substances from logical is a thought-provoking job in which we face numerous practical/utilization things, for instance, Biological system, bioinformatics and biomolecules including (DNA, RNA). (NER) is grounded on machine learning, ordinarily; it utilized Machine Learning for NER which are statistical methods. For example BN, Naive Bayesian (NB) and for rule based, for instance, Conditional Random Fields and so on. Here, in this portion, we supply a general summary of a statistical methods or techniques which are utilized for Named Entity Recognition that study carries through [6]. The name of individual, association, area and so on was discovered with the utilization of SS algorithm CRF for Named Entity Recognition. The framework or system has revealed, in which the accuracy is very close to the human level. In [7] Maximum Entropy Classifier is utilized for biomedical Named Entity Recognition. The proposed System utilizes GENIA corpus to characterize and recognize the numerous biomedical nomenclature or taxonomy, for instance, DNA, Proteins, types of Cell, RNA as well as other bodily structures. Due to the anatomical figure/construction as well, an overview of content which belongs this, it is tough to compact with complete accuracy higher than 80.00% for Machine Learning method. For the very reason, that is, Morphological and spelling variation of bio medical substances, probabilities categorized in numerous groups. Henceforth, an enhanced Set is needed for Biomedical Named Entity Recognition to adjust to these problems feature set, as; Affixes, Orthographic, Uni-grams be presented and represented [1]. It joined high dimensional characteristics for Biomedical Named Entity Recognition with the utilizing of multi cast Support Vector Machine.

The Biomedical Named Entity Recognition which assures that any substances hold an alternate substance in them which located in its bounds is mentioned as nested entity. Conditional Random Fields (CRF) which is broadly utilized for named entity recognition as well is beneficial for discrimination/identifying something of nested named entities. According to [4], [8] a methodology which is identified as discriminating constituency parser is recommended to execute nested NER by transmutation or change each phase or term into tree and such methodology which implemented to daily paper, bio-medical work, the outcomes were more precise as compared traditional SS Conditional Random Field.

Bio medical Named Entity Recognition has likewise expanded the enthusiasm of discovering illness names in online content, various works for the vindication of cancer disease are available on the web, to let them free to users to utilize these numerous tools and techniques for tumor treatment. For prescription different clinical notes or records have been examined by the specialists and researchers and according to that investigated report, the combination of Support Vector Machine (SVM) and Conditional Random Fields succeeded well execution in analysis in the field of medical mining, with utilizing the similar data set utilized as a part [3]. Additional study or examination has been completed on doctor's facility release synopses. Further, much more characteristics precision of the framework has been expanded [10] in that framework features are morphology, orthographic, semantic tags and so on features. The Respiratory disease is most normal ailment and there are numerous medicinal drugs or pharmaceutical available for its cure, with a specific end goal that a collection of facts study/examination has been completed and it proves the latent worth in text mining in the field of respiration medicine [11]. In Expanding exploration or survey of various data source including protein, gene as well Bio Tagger was prepared on it in the domain of medical text mining. The Experimentation or Testing result demonstrated in which Bio Tagger conceivably valuable to extract the protein, gene in the form of huge dataset accommodated for the Training [12]. Content characteristics dependably assume a vital part in Named Entity Recognition; the framework's execution can be considerably enhanced via expansion of many characteristics. In [5] dictionary based characteristics have been utilized because of ailment Named Entity Recognition it made through choosing the low accuracy and high recall, it expels loud terms. And utilizing these characteristics Support vector Machine was prepared and the outcomes got 11.3% which more precise as compared to the former/old strategy or technique.

III. Proposed Method

In this part, we give the explanation in details of feature selection, classification scheme and proposed classifier fusion method.

A. Feature Extraction and Selection

To build a classification model, feature selection takes an important role in data classification. In this research, our utilized feature set is based on local feature and non-local features. In this regards, we extract local features from token whereas, the non-local characteristics relies upon local feature - POS tags, sliding window feature, and so on. The detailed information of this section is divided into below subsections.

1) Orthographic Features: Geometry and indentation of the text, for instance, digits, numeric, numbers, capitalization, single cap, two caps, all caps, symbols, punctuation and etc; these kinds of features are very efficient in Named Entity Recognition. In past few researches, the use of orthographic features is widely advocated in [12]-[14]. Our used experimental orthographic features are shown in Table I.

TABLE I. List of Orthographic Features

Features	**Examples**
Upper Case	DGS, EMD,AT,NKH, ALD
Hyphens	X-ALD, X-linked, dopa-responsive
Alphanumeric	DFNB4,SCA6,G6PD

2) Part of Speech Tags: POS Tags is supportive, to identify the boundary of words. With specific cases, the author

shows better performance in part of speech tags [15]. Whereas its well-established certainty that tagging Part of Speech is hard as well rich computing procedure, therefore investigators or scholars have precluded that the utilizing of Part Of Speech Tagging because of its limited performance of named entity recognition [10], [16]. Our includes NEs and contextual features for POS tagging.

3) N-Grams: N-Grams, It is the model and fundamentally a framework of linguistic/language and it grounded on the principles of grammar. N-gram grounded rules that well portrayal of words has better execution of data recovery. Normally utilized phased or content mixtures are unigram, it produces an entire sentence in a one set or pair, and the others, bi-gram and tri-gram combination are used which are high dimensional. In general, N-gram are expressed via question,

$$P(W) = \prod_{i=1}^{|W+1|} P(w_i|w_0 \dots\dots\dots w_{i-1}), \quad (1)$$

In N-grams, the representation of uni-grams is $P(w_i|w_0 \dots\dots\dots w_{i-1}) \approx P(w_i)$ as equation (1), for bi-grams we put or add one portion in the first equation of uni-grams we found the equation of bi-grams which can be denoted, $(w_i|w_0 \dots\dots\dots w_{i-1}) \approx P(w_i|w_{i-1})$. Here in this experiment just uni-grams and b-grams has been incorporated. Therefore, through this method we can find tri-grams and as well other N-grams models too.

4) Affixes: The prefix and suffix features always show considerable performance within named entity recognition. In this regards, few researchers have proposed the utilization of named entity in their own particular way. The authors [13], [17] has gathered most common prefix and suffix from training data. Whilst [12], [18] the author gathered 23 categories of prefix and suffix data using statistical methods as their own distribution. Our experiment shows the significant improvement in contextual features affixes. In our experiment prefix and suffix which created in such method for instance. “Adenomatous polyposis coli tumor” signifies the designation of the illness. Such as prefix and suffix development and the two characters has been occupied from every term and henceforth the prefix built is “adpocotu” and the suffix framed "usislior” respectively.

5) Contextual Features: It alludes to the word going before and pursuing the named elements, e.g. (named element), so for each element, we utilize two token cases about this, for example, $c = (w_{-2}, w_{-1}, w_0, w_1, w_2)$ currently for every token w_0 it shows up under that area $w_i, w_{i+1}, w_{i+2} \dots\dots\dots w_n$ and according to the second equation named as contextual window, C= $\prod_{i=-2}^{2} w_i$ via this you can compute more particular as well as similar characteristics. In our test contextual characteristics are the much more vital features in the Named Entity Recognition joined with the affixes. At first two contextual features took after by the present word were chosen for the analysis, yet understanding the significance of these features four contextual characteristics as appeared in (2) has been chosen. The blending of both two contextual and affixes features has demonstrated the well precision instead of other features. And both two are, in this the arguments of two words which happens before and as well two happens after in the named entities.

B. Classification Scheme

According to this literal composition, it totally shows that Machine Learning Method concentrated for NER. For this experiment; from Rule Based Learners such as Partial Decision Trees, Non-Nested Generalized Exemplars and Naive Bayesian Decision Table and supervised a set of Machine Learning Methods as, Naive Bayesian and Bayesian Network has been preferred. Further, the characterization plans get from this area. The Prevalent Data Mining tool broadly utilized by researchers and professors named as WEKA, and in this experiment classifiers utilized as a part of this experiment use up from WEKA [19], [20]. And the selected classification scheme accomplished a considerable execution by utilizing the National Center for Biotechnology Information (NCBI) Training Dataset by 10 Fold cross validation.

1) Bayesian Network (BN): Bayesian Network generally utilized for content classification [13] and it is supervised parametric classifier. Bayesian systems, beginning from Bayesian hypothesis and it is the kind of systems which is made of the set out of nodes represented by U, U= $\{X_1, X_2, X_3 \dots\dots X_n\}$. These nodes are reticulated amongst another through an arrow set indicated through A, where A depends upon set of principles and characterized as, A= $\{(X_i, X_j)|X_i, X_j \in U, i \neq j$ [8]. Consequently if there is a connection between nodes then they ought to rely upon each different as expressed by the Bayesian hypothesis, the connection amongst nodes denoted via an arrow. An arrow from node Y to node X signifies that Y node is the parent of node X. According to Bayesian network child node must, be autonomous of parent node or fulfill the Markov Condition. As hypothesis $P(X_1, X_2, X_3 \dots\dots X_n)$ would therefore stay able to be established as demonstrated as follows:

$$P(x_1, x_2, x_3 \dots\dots x_N) = \prod_{i=1}^{N} P(x_i|pa(X_i)), \quad (2)$$

The formula which mentioned above in that formula or equation Parent variable shows via pa. Execution of Bayesian Network were assessed on Training Dataset utilizing 10 fold cross validation, comes about on joining every one of the characteristics indicated accuracy of 0.872%, Recall of 0.833% and F-score of 0.844% which appeared in Table III. However, the combination of Affixes and Contextual features has been accomplished the F-sore of 0.861%.

2) Naive Bayesian (NB): The Naive Bayesian, which has its starting point from Bayes hypothesis as well-known as a probabilistic supervised classifier. Notwithstanding Bayesian hypothesis presumption is included and henceforth each prospect is considered freely toward a basic leadership. The straightforwardness and simplicity of preparing of Bayesian make it perfect for complex order issues [19]. Since accepting each element to be autonomous of each other so as opposed to computing the variance of an individual element, co variance matrix is created [9]. Mathematically Bayesian,

$$P(C|x_1 \dots \dots x_n) = \frac{p(C)p(x_{1\dots\dots}x_n|C)}{p(x_{1\dots\dots}x_n)}, \quad (3)$$

The features in this formula or equation $x_1 \dots \dots x_n$ self-sufficient of the class as well each other and the C in the equation indicate the class. With utilizing the Naive Bayesian outcomes got is the F-sore of 0.858% on every one of the features consolidated. As like BNs have been seen here, affixes feature and contextual feature has been accomplished the F-score of 0.870% which smashed the execution of all characteristics joined.

3) Naive Bayesian Decision Table (DTNB): DTNB, actually it is a semi NB method which joined decision table and after joined the better precision has demonstrated by the Naive Bayesian Decision Table as compared previous Naive Bayesian. The amalgamation of two methods NB and decision table generates a network system, in that network the decision table symbolizes probability table and this network system considerably parallel to BN. In our case, Naive Bayesian Decision Table has shown better outcomes as compared to NB and BN. The Parameters/Limitations for Naive Bayesian Decision Table were presented as; cross validation value is set to '1', display Instructions is set to 'False', utilize IBK is set to 'False' and look is instated with In reverse with erase. DTNB has accomplished better outcomes contrasted with the general classification scheme; it has beaten methods like Bayesian Network, Naïve Bayesian, Partial Decision Trees and Non-Nested Generalized Exemplars. The Combination of affixes feature, orthographic feature, affixes feature and N-gram feature has been accomplished the best F-score of 0.874% whereas F-score of 0.872% by contextual and affixes.

4) Non-Nested Generalized Exemplars (NNGE): In 1995 by Bent this Non-Nested Generalized Exemplars were firstly introduced, Generalization completed utilizing blending the models to frame hyper rectangle which presents conjunctive rules with interior dis-junction [11], [21]. NNGE has demonstrated better precision [19], at whatever point another example is added to the dataset of training the classifier performs hypothesis through the connection the recent example of the Closest Neighbor of that class. Various endeavors to attempts the hypothesis is set to 5 and the endeavors of the fold for mutual information are also introduced with 5. The grouping of affixes and contextual has been accomplished the best F-score of 0.865% whereas the F-score of 0.841% has acquired by all features joined.

5) Partial Decision Trees (PART): With thee consolidating C4.5 and RIPPER and subsequently is capable rule based learner. The merit of Partial Decision Trees above RIPPER is its straightforwardness since it over and over produces PART as opposed to the intricate progress phases took after by RIPPER [5], [22]. Parameters of Partial Decision Trees are instated as twofold part is set to false. After joining Contextual feature, Orthographic feature and Affixes Feature Partial Decision Trees accomplishes the F-score of 0.723% and partial decision trees is the main classifier which has demonstrated poor execution in this challenge. Though when contextual, Affixes, Orthographic and N-grams are provided as features at that time PART execution is the most noticeably awful and accomplishes F-score of 0.537%.

C. Classifier Fusion

This technique is acquainted with enhancing the exactness above single classifier and creates the execution livelier vigorous in contradiction of every distinct method. Joining method acquires the attributes of the different order conspire and thus a capable group is created. Methods or techniques are consolidated in light of normal probabilities. In normal of probabilities, the likelihood can be accomplished as, $\widehat{P} = \frac{1}{L}\sum_{j=1}^{L} P_j$ whereas P_e represents error probabilities and computed via $P_e = \Phi\left(\frac{\sqrt{L}(0.5-P)}{\sigma}\right)$ and $P_1, P_2 \dots \dots P_L$ Are free or independent probabilities [12]. Inside and out an examination of order match has been completed which extend from two pairs combinatory to five pairs combinatory or blend. Combination of classifier has been done utilizing Vote in WEKA. At first, we utilized training dataset in the test, and 10 Fold cross validation has been connected. Right off the bat or initially, we consolidate two sets of classifiers. At that point, we joined three, four and five sets classifiers individually. The outcomes in the subtle elements appear in the following segment.

IV. Experimental Method

A. Data Set

The National Center for Biotechnology Information (NCBI) ailment corpus which is unreservedly accessible by NCBI on which this test or experiment is based. The corpus incorporates 793 synopses compositions which comprise of 2783 sentences and an aggregate of 6900 malady names [13]. Contrasted with AZDC corpus NCBI corpus contains 3224 one of a kind infection names [5]. Explanations were finished utilizing a web base device called PubTator [13], [23]. Table II cited from (NCBI) which shows list of Data set features we have utilized as a part of our test.

The corpus comment was relegated four classifications in view of the idea of the sickness which comprises of 3922 particular illness explanation, 1029 malady family or category explanation, 173 complex and 1774 modifier notices. Additionally, the dataset is isolated within Training Set, Testing Set and Development Set.

As of Table III persuaded presumption can be prepared, initially, we saw the distinct methods which indicated bad execution, for example, Bayesian Network, Naive Bayesian, Partial Decision Trees and Non-Nested Generalized Exemplars contrasted with Naive Bayesian Decision Table. Meanwhile Naive Bayesian Decision Table is a mixture method which joins Decision Trees and Naive Bayesian, also its guaranteed that completely list of capabilities, for example, orthographic, N-grams and Part Of Speech tags are not valuable in the acknowledgment of Biomedical disorder names, in practically each event it has been seen in which affixes and contextual have accomplished well outcomes.

TABLE II. USED DATASET

Classes	Training Set	Testing Set	Dev. Set
Modifiers	1292	264	218
Specific Disease	2959	556	409
Composite Mention	116	20	37
Disease Class	781	121	127

TABLE III. 10 FOLD CROSS VALIDATION ON AVAILABLE FEATURE SET

Classifier	Features	P	R	F
Bayesian Network	Contextual	0.838	0.848	0.848
	Contextual +Affixes	0.870	0.855	0.868
	Contextual +Affixes+N-grams	0.866	0.828	0.84
	Contextual +Affixes+POS Tags	0.869	0.839	0.847
	Contextual +Affixes+Orthographic+N-grams	0.874	0.828	0.843
	Contextual +Affixes+Orthographic+ N-grams +POS Tags	0.872	0.833	0.844
Naive Bayesian	Contextual	0.827	0.845	0.831
	Contextual +Affixes	0.873	0.873	0.870
	Contextual +Affixes+N-grams	0.857	0.851	0.852
	Contextual +Affixes+POS Tags	0.865	0.858	0.859
	Contextual+Affixes+Orthographic +N-grams	0.865	0.851	0.856
	Contextual+Affixes+Orthographic +N-grams+POS Tags	0.868	0.854	0.858
Decision Table Naïve Bayesian	Contextual	0.840	0.852	0.842
	Contextual+Affixes	0.875	0.876	0.872
	Contextual +Affixes+N-grams	0.876	0.873	0.871
	Contextual+Affixes+POS Tags	0.869	0.868	0.866
	Contextual+Affixes+Orthographic +N-grams	0.875	0.876	0.874
	Orthographic+N-grams+POS Tags	0.871	0.874	0.872
Non-Nested Generalized Exemplars	Contextual	0.848	0.845	0.841
	Contextual+Affixes	0.868	0.869	0.865
	Contextual+Affixes + N-grams	0.846	0.847	0.841
	Contextual+Affixes +POS Tags	0.846	0.847	0.841
	Contextual+Affixes + Orthographic+N-grams	0.846	0.847	0.841
	Contextual+Affixes + Orthographic+N-grams+POS Tags	0.846	0.847	0.841
Partial Decision Trees	Contextual	0.779	0.773	0.668
	Contextual+Affixes	0.768	0.736	0.693
	Contextual+Affixes + N-grams	0.747	0.631	0.528
	Contextual+Affixes +POS Tags	0.757	0.685	0.616
	Contextual+Affixes+ Orthographic+N-grams	0.748	0.636	0.537
	Contextual+Affixes+ Orthographic+N-grams+POS Tags	0.758	0.687	0.619

Promote research has been completed on the designated characteristics in other words "Affixes and Contextual" for the arrangement. Also, we have investigated mix of methods to enhance the outcomes. Hence we have consolidated distinctive classifiers.

B. Baseline Method

We have compared our method with BANNER Bio-Medical Named Entity Recognition [5].

As of Table IV, it is clear that the most elevated F-score has been accounted for by the blend of both NB as well Naive Bayesian Decision Table it revealed most astounding F-score of 0.876 and accuracy of 0.878. Though the least F-score has been accounted for by the compound of Naive Bayesian and Non-Nested Generalized Exemplars, it acquired 0.865 of F-score. As of Table IV, unmistakably mix of two sets of classifier has beaten the single order comes about. Contrasting the consequences of Tables IV and III we have discovered that enhanced outcomes have been accounted for via two sets combination of classifiers. In addition, the investigation has been completed and three sets of classifiers have been joined and the outcomes showed in Table V.

C. Results and Discussions

Fascinating outcomes has been gotten within Table V. As of Table IV, top F-score were accounted for via compound of Naive Bayesian and Naive Bayesian Decision Table whilst the most reduced F-score were accounted for via Naive Bayesian and Non-Nested Generalized Exemplars. Within the table, most reduced F-score has been accounted for through mix of Naive Bayesian+ Naive Bayesian Decision Table consist of 0.874% whilst most elevated F-score has been accounted for through mix of Naive Bayesian+Bayesian Network+Non-Nested Generalized Exemplars and accomplished 0.885% of F-score. This reality is on account of Non-Nested GE executes close neighbor like algorithm and consequently utilizing three distinctive methods whilst BN, Decision Table and Naive Bayesian Decision Table, demonstrated worse execution is because of DT from the time when Naive Bayesian Decision Table shapes Hybrid Naive Bayes and thus Partial Decision Trees while joined through Bayesian Network and Naive Bayesian is not fit for enhancing execution. Looking at the after effect of Naive Bayesian Decision Table + Naive Bayesian through Naive Bayesian+Bayesian, Network+Non-Nested Generalized Exemplars important change has been noticed.

As of Table V unmistakably mix of three sets of techniques has beaten the consequences of two sets of methods. It provides the inspiration for further joining four sets of classifiers. Moreover, a blend of methods has been accounted for in Table VI.

TABLE IV. COMBINATION OF TWO PAIRS CLASSIFIER

Classifier	P	R	F
BN + NB	0.875	0.876	0.872
BN + DTNB	0.878	0.879	0.877
BN + NNGe	0.867	0.869	0.865
BN + PART	0.878	0.880	0.878
NB + DTNB	0.877	0.88	0.875
NB + NNGe	0.867	0.869	0.865
NB + PART	0.873	0.875	0.870
DTNB + NNGe	0.867	0.869	0.865
DTNB + PART	0.866	0.864	0.853
NNGe + PART	0.868	0.869	0.865

TABLE V. COMBINATION OF THREE PAIRS CLASSIFIERS

Classifier	P	R	F
BN+NB+DTNB	0.880	0.881	0.879
BN+NB+NNGe	0.884	0.885	0.882
BN+NB+PART	0.876	0.878	0.875
BN+DTNB+NNGe	0.889	0.89	0.887
BN+DTNB+PART	0.881	0.884	0.88
BN+DTNB+NNGe	0.889	0.89	0.890
NB+DTNB+NNGe	0.889	0.889	0.884
NB+DTNB+PART	0.877	0.878	0.871
NB+NNGe+PART	0.881	0.884	0.880
DTNB+NNGe+PART	0.884	0.883	0.876

TABLE VI. COMBINATION OF FOUR AND FIVE PAIRS CLASSIFIERS

Classifier	P	R	F
BN+NB+DTNB+NNGe	0.884	0.885	0.883
BN+NB+NNGe+PART	0.879	0.882	0.878
BN+NB+NNGe+PART	0.890	0.888	0.887
BN+DTNB+NNGe+PART	0.888	0.888	0.883
NB+DTNB+NNGe+PART	0.890	0.889	0.883
BN+NB+DTNB+NNGe	0.884	0.885	0.3883
BN+NB+DTNB+PART	0.879	0.882	0.878
BN+NB+NNGe+PART	0.890	0.890	0.890
BN+DTNB+NNGe+PART	0.888	0.886	0.883
NB+DTNB+NNGe+PART	0.890	0.889	0.883
BN+NB+DTNB+NNGe+PART	0.890	0.880	0.887

Table VI speaks to a combination of four and five sets of methods. As of Table VI it is seen that mix of Naive Bayesian+Bayesian, Network+Non-Nested, Generalized Exemplars+ Naive Bayesian Decision Table has demonstrated the most minimal execution contrasted with a mix of NB, BN, NNGE and PART while, in addition, we have seen that the union of five classifiers which demonstrated the output and according to that generally no change/enhancement of F-score, Accuracy and Recall as well. The comparing of the outcomes which are achieved via Table VI with the Table V and Table IV as well as through that achieved outcomes we have seen the vital enhancement discovered in the Accuracy, Recall and F-score.

Contrasting single sets of a classifier which examined, it states that 87.4% of F-score accomplished by Naive Bayesian Decision Table on characteristics, for example, affixes, contextual, orthographic and N-grams. More of a thing detected that union of Naive Bayesian and Naive Bayesian Decision Table accomplished 87.6% of F-score and compacted separate order consequence of Naive Bayesian Decision Table. For example, with utilizing the contextual and affixes characteristics appeared within Table IV. The union of three sets of methods has come out better with the past outcome as well utilized same characteristics; fusion of Naive Bayesian+Bayesian, Network+Non-Nested, Generalized Exemplars accomplished the 88.5% of F-score. Whilst the 88.7% of F-score with the utilizing the fusion of four sets methods as, Naive Bayesian+Bayesian, Network+Partial Decision, Trees+Non-Nested, Generalized Exemplars and it indicated the outperformed outcomes.

In Fig. 1, it seems that the grouping of two rule based (NNGE, PART, DTNB) and statistical methods (BN, NB) gave a better outcome, and in Fig. 1 the examination of various union pairs has been completed. In General, it has been going that, union of four sets classifiers has present better outcomes as compared with the union of three, two and single/one pair(s) of classifiers and accomplished a totally precision on training dataset is 89%.

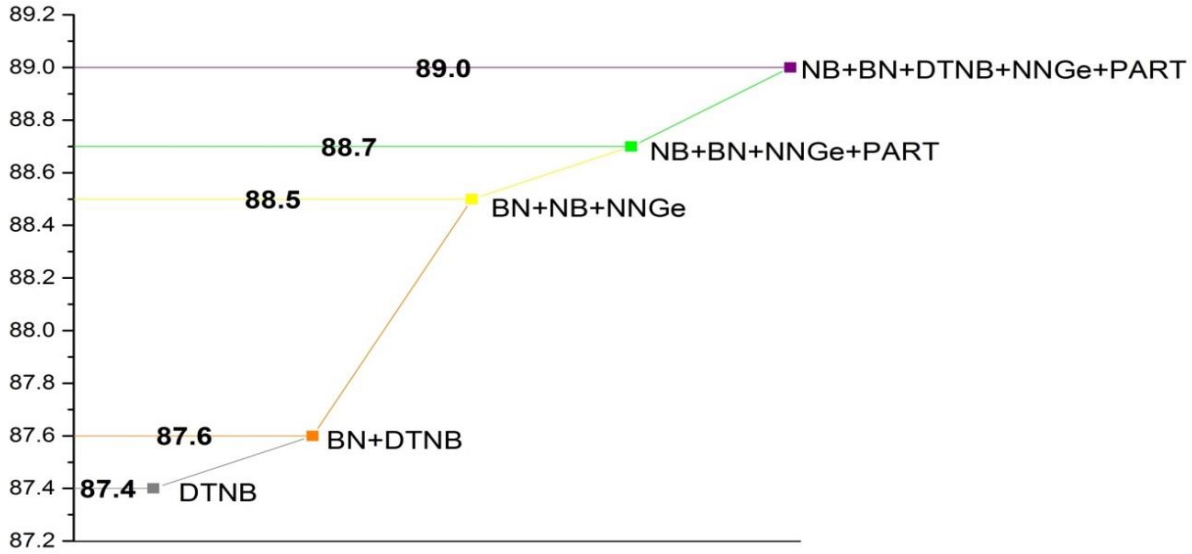

Fig. 1. Overall accuracy by available classifiers.

Moreover, we broadened our approach and the connected combination of four sets of classifiers utilizing affixes and contextual characteristics on testing and developing a set. Table VII demonstrates the consequences of applying the combination on three distinct datasets viz., Training Set, Testing Set and Developing Set. On Training set, 10 Fold cross validation has been implemented whilst, whatever remains of Datasets, Training has been finished on the Training Dataset and Testing has been passed on Testing dataset and development dataset, comes about on these datasets has appeared within Table VII.

Table VII demonstrates that the outcomes acquired on Training Set, Testing Set and Development Set is via fusion technique. In addition, these are the values or Results (F-score) on these sets via fusion technique is like, on Training set 88.7% of F-score, on Testing set 86.4% of F-score, though on Developing set 83.5% of F-score has been analyzed. Our outcome has been contrasted with the benchmark system [13]. Extensively, and for longer period, this has been demonstrated that union of fusion classifier method is the finest method for Disease/Illness NER.

According to Fig. 2, it is showing that the outcomes were acquired via Propose Method after the comparison between Proposed Method and BANNER Method. Finally Proposed Method had beaten the BANNER Method [5] outcomes. On Training set 84.5% of F-score, on the Testing set 81.8% of F-score and Development set 81.9% of F-score and it is presented within Table VII.

TABLE VII. COMBINATION OF FOUR AND FIVE PAIRS CLASSIFIERS

System	Dataset	P	R	F
Proposed Result	**Training**	**0.890**	**0.890**	**0.890**
	Testing	**0.870**	**0.866**	**0.864**
	Development	**0.840**	**0.841**	**0.835**
BANNER Result	Training	0.867	0.826	0.845
	Testing	0.838	0.800	0.818
	Development	0.821	0.818	0.819

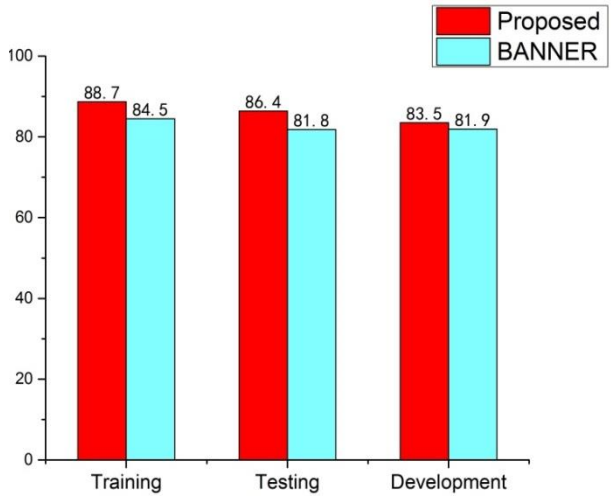

Fig. 2. Proposed method compared with BANNER

V. CONCLUSION

This research paper is aimed at bio-Medical Named Entity by proposing the approach of Hybrid Machine Learning. The performances of different approaches viz., Machine learners like, Naïve Bayesian, Rule Based Learners i.e. PART, DTNB and NNGE, and Bayesian Network, are compared. Investigation and exploration of the data discovers that execution close to the best in class can be accomplished via a blend of Statistical Machine Learning and Rule Based Techniques utilizing straightforward characteristics such as; contextual and affixes. Amalgamation of four sets i.e. (NB, BN, PART and NNGE) has accomplished overall precision on Training dataset, Development dataset and Testing dataset with 89.0%, 84.0% and 86.0%, respectively. This Classifiers blending of two, three, four and five has been utilized to investigate the execution of sets of classifiers via vote WEKA Data Mining Tool. The standard BANNER results are outperformed by this fusion approach which has given far better results on the same dataset. In the future we will apply and check the effectiveness of our proposed method for Drug Name Recognition.

REFERENCES

[1] Habib, M.S. Addressing scalability issues of named entity recognition using multi-class support vector machines. World Academy of Science, Engineering and Technology.–2008.–37.–P 2008, 69-78.

[2] Huang, Z.; Hu, X. Disease named entity recognition by machine learning using semantic type of metathesaurus. International Journal of Machine Learning and Computing 2013, 3, 494.

[3] Patrick, J.; Li, M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. Journal of the American Medical Informatics Association 2010, 17, 524-527.

[4] Li, L.; Fan, W.; Huang, D. A two-phase bio-ner system based on integrated classifiers and multiagent strategy. IEEE/ACM transactions on computational biology and bioinformatics 2013, 10, 897-904.

[5] Leaman, R.; Gonzalez, G. In Banner: An executable survey of advances in biomedical named entity recognition, Pacific symposium on biocomputing, 2008; Big Island, Hawaii: pp 652-663.

[6] Liao, W.; Veeramachaneni, S. In A simple semi-supervised algorithm for named entity recognition, Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing, 2009; Association for Computational Linguistics: pp 58-65.

[7] Patrick, J.; Wang, Y. In Biomedical named entity recognition system, Proceedings of the Tenth Australasian Document Computing Symposium (ADCS 2005), 2005.

[8] Finkel, J.R.; Manning, C.D. In Nested named entity recognition, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, 2009; Association for Computational Linguistics: pp 141-150.

[9] Zhu, F.; Patumcharoenpol, P.; Zhang, C.; Yang, Y.; Chan, J.; Meechai, A.; Vongsangnak, W.; Shen, B. Biomedical text mining and its applications in cancer research. Journal of biomedical informatics 2013, 46, 200-211.

[10] Doan, S.; Xu, H. In Recognizing medication related entities in hospital discharge summaries using support vector machine, Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010; Association for Computational Linguistics: pp 259-266.

[11] Piedra, D.; Ferrer, A.; Gea, J. Text mining and medicine: Usefulness in respiratory diseases. Archivos de Bronconeumología (English Edition) 2014, 50, 113-119.

[12] Torii, M.; Wagholikar, K.; Liu, H. Using machine learning for concept extraction on clinical documents from multiple data sources. Journal of the American Medical Informatics Association 2011, 18, 580-587.

[13] Doğan, R.I.; Lu, Z. In An improved corpus of disease mentions in pubmed citations, Proceedings of the 2012 workshop on biomedical natural language processing, 2012; Association for Computational Linguistics: pp 91-99.

[14] Collier, N.; Takeuchi, K. Comparison of character-level and part of speech features for name recognition in biomedical texts. Journal of Biomedical Informatics 2004, 37, 423-435.

[15] Shen, D.; Zhang, J.; Zhou, G.; Su, J.; Tan, C.-L. In Effective adaptation of a hidden markov model-based named entity recognizer for biomedical domain, Proceedings of the ACL 2003 workshop on Natural language processing in biomedicine-Volume 13, 2003; Association for Computational Linguistics: pp 49-56.

[16] Ratinov, L.; Roth, D. In Design challenges and misconceptions in named entity recognition, Proceedings of the Thirteenth Conference on Computational Natural Language Learning, 2009; Association for Computational Linguistics: pp 147-155.

[17] Kazama, J.i.; Makino, T.; Ohta, Y.; Tsujii, J.i. In Tuning support vector machines for biomedical named entity recognition, Proceedings of the ACL-02 workshop on Natural language processing in the biomedical domain-Volume 3, 2002; Association for Computational Linguistics: pp 1-8.

[18] Zhou, G.; Su, J. In Named entity recognition using an hmm-based chunk tagger, proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002; Association for Computational Linguistics: pp 473-480.

[19] Neelakantan, A.; Collins, M. Learning dictionaries for named entity recognition using minimal supervision. arXiv preprint arXiv:1504.06650 2015.

[20] Frank, E. Fully supervised training of gaussian radial basis function networks in weka. 2014.

[21] Roy, S. Nearest neighbor with generalization. Christchurch, New Zealand 2002.

[22] Frank, E.; Witten, I.H. Generating accurate rule sets without global optimization. 1998.

[23] Wei, C.-H.; Kao, H.-Y.; Lu, Z. In Pubtator: A pubmed-like interactive curation system for document triage and literature curation, Proceedings of the BioCreative 2012 Workshop, Washington, DC, 2012; pp 20-24.