Investigation of Minimally Invasive Robot-assisted Brain Surgery Using Automatic Speech Recognition

Dr.Jagadeesh Kumar[1] and Dr.Susan Daenke[2]

[1]Marik Institute of Computing, Artificial Intelligence, and Machine Learning, NIMS Institute of Engineering and Technology, NIMS University Rajasthan, Jaipur - 303121, India.

[2]Department of Medical Sciences, University of Oxford, Oxford - OX3 9DU, United Kingdom.

Abstract: **T**his paper philanthropy the on-going machine learning as utilized in the cutting edge and as relevant to imminent automatic speech recognition strategies on minimally invasive robot-assisted brain surgery. The aspiration is to advance experimental cleansing among the machine learning and automatic speech recognition groups that have unfolded in the point of reference of surgical robots in health care. The inventive facsimiles are organized by the central machine learning archetypes that are besides vogue at this point or have indolence for construction that shatter hand-outs to automatic speech recognition aptitude. The measures offered and tangled in this manuscript comprehend: adaptive and multi-task learning, active learning, bayesian learning, discriminative learning, generative learning, supervised and unsupervised learning. The learning exemplars are presented in the point of view of automatic speech recognition instruments and capacities in biomedical robots for surgery. This structure passes on and reviews topical advances of deep learning and learning with machine intelligence, further spotlight is on their ceaseless importance in the development of hybrid deep learning based automatic speech recognition in the perspective view of robot-assisted surgeries.

Keywords: Automatic Speech Recognition, Biomedical and Surgical Robots, Deep Learning, Health Care, Machine Intelligence

1 Introduction

The major aspiration of this manuscript is to propose insight from diverse stance while systematizing many Automatic Speech Recognition (ASR) techniques into an entrenched Machine Learning (ML) performance (M.Dhanalakshmi, T.Nagarajan, P.Vijayalakshmi, 2018). Further explicitly, this article offers a general idea of broad ASR methods by creating numerous means of organizing and classifying the frequent ML archetypes, functioned by their learning fashion. The learning manners upon the classification of the learning methods that are referred to the key characteristics of the ML algorithms, like the feature of the algorithm's input or output, the decision function exercised to establish the classification or identification of output, and the loss function utilized in training the interpretations. Whereas convoluting on the key discerning factors related with the dissimilar groups of the ML algorithms, unique deliberation is compensated to the associated sculpture urbanized in ASR explore (D.Reynolds, 2002). In its broad extent, the aspire of ML is to build up automatic schemes proficient of simplifying from beforehand observed patterns, and it realizes so by creating or learning practical reliance amid the subjective input and output realms. ASR, which is anticipated to translate the audio information in speech statistics into its fundamental linguistic pattern, classically in the form of speech series, is therefore essentially an ML predicament; i.e., specified instances of influences as the incessant audio characteristic series or perhaps sound waves and outputs as the ostensible valued label series (declaration, telephone, or expression), the objective are to envisage the latest production series from an original input string (E.Nöth, S.Harbeck, and H.Niemann, 1999). This prophecy assignment is frequently termed as categorization when the sequential section restrictions of the output labels are understood as known. If not, the prophecy task is called identification. For instance, phonetic classification and phonetic identification are two miscellaneous errands: the earlier with the phone restrictions given in both training and testing data, whilst the latter entails no such periphery data and are therefore further intricate (T.Schultz, A.Waibel, 2001). Correspondingly, secluded idiom "identification" is a typical taxonomy aspect in ML, apart from with a random aspect in the input space owing to the capricious length of the speech contribution. Moreover, continuous speech recognition is an unique kind of structured ML tribulations, where the prophecy must persuade further constrictions with the output having structure.

The utilization of robotics in therapeutic wellbeing isn't new. Robots as of now help out in spinal medical procedure, with models, for example, Renaissance enabling specialists to tighten spines with 99 percent precision (9 percent higher than traditional methods). The popular da Vinci careful framework (where specialist's hand movements are adapted into littler, increasingly exact mechanical developments) is currently utilized over a wide scope of dealings, from prostate malignant growth treatment to performing heart valve medical procedure. In the US, a robot called Watson aids conclusions and produces the executives' plans for oncology patients by blending data from a large number of reports, understanding records, clinical preliminaries and diaries (S.Furui, 1991). In the interim, Woebot, the world's first mechanical therapist, has in excess of two million discussions every week. Despite the fact that experts at the Children's National Medical Center in Washington have as of late built up a careful robot (called STAR) which can suture delicate tissue. The criticism is that still have far to go before a patent minimal effort with enough adroitness and affectability in robots are worked to play out the sort of work talked about.

2 Background

Think about the canonical setting of regression in AI. Expect a preparation set; apprehended from the propagation *p(x, y), x∊X, y∊Y*.

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{m}$$

The objective of learning is to discover a choice capacity $f: X \rightarrow Y$ that accurately predicts the yield of a future info drawn from a similar dissemination. The forecast errand is called grouping when the yield takes straight out qualities, which is expected in this work (R.Rojas, 1996). In a multi-class setting, a choice capacity is dictated by a lot of discriminant capacities, i.e.

$$f(x) = arg_y ma\,x\, d_y(x) \qquad (1)$$

Each function d_y is a class-subordinate capacity of x. In two-fold characterization where $Y = \{\pm 1\}$, be that as it may, usually to utilize a solitary "discriminant work" as pursues,

$$f(x) = sgn\, d(x) \qquad (2)$$

Ritualistically, learning is worried about finding a choice capacity (or identically a lot of discriminant capacities) that limits the normal hazard, i.e.

$$R_p(f) = E_p(x, y)[L(f(x), y)] \qquad (3)$$

where L (f(x), y) is the misfortune work. The misfortune work measures the "cost" of settling on the choice $f(x)$ while the genuine yield is y; and the normal hazard is just the normal estimation of such an expense. In ML, it is essential to comprehend the contrast between the choice capacity and the misfortune work. The previous is regularly alluded to as the "model". For instance, a straight model is a specific type of the choice capacity, implying that input highlights are directly joined at order time (V.A.Petrushin, 2000). Then again, how the parameters of a direct model are assessed relies upon the misfortune work (or, proportionally, the training objective). A specific model can be judged utilizing diverse adversity capacities, although a similar adversity capacity can be associated to an assortment of models. Unmistakably, the normal hazard is difficult to upgrade straightforwardly as *p(x,y)* is commonly obscure. By and by, it is regularly meant to discover a choice capacity that limits the exact hazard, i.e.

$$R_{emp}(f) = \frac{1}{m}\sum_{i=1}^{m} L(f(x^{(i)}), y^{(i)}) \qquad (4)$$

with respect to the preparation set. It has been appeared, if L fulfills certain imperatives, $R_{emp}(f)$ meets to $R_p(f)$ in likelihood. The preparation set, notwithstanding, is quite often lacking. It is in this way significant to apply particular sort of regularization to improve speculation. This prompts a down to earth preparing objective alluded to as exactness regularization which takes the accompanying general structure:

$$J\,(f) = R_{emp}(f) + \gamma\, C(f) \qquad (5)$$

where $C(f)$ is a regularizer that measures "multifaceted nature" of f, and λ is a tradeoff parameter. Truth be told, a key issue in ML is to infer such types of that ensure the speculation implementation of learning. Among the most prevalent hypotheses on speculation blunder bound is the VC dimension hypothesis. As indicated by the hypothesis, if two models portray the preparation information similarly well, the model with the littlest VC measurement has better speculation execution (E.Douglas-Cowie, N.Campbell, R.Cowie, and P. Roach, 2003). The VC measurement, in this way, can customarily fill in as a regularizer in meticulous hazard minimization, gave that it has a logically cooperative structure, as on account of expansive edge hyper-planes. On the other hand, regularization can be seen from a Bayesian viewpoint, where it is regarded as an irregular variable. One needs to determine an earlier opinion, indicated as $p(f)$, before observing the preparation information (D). Interestingly, the back likelihood of the model is inferred in the wake of preparing information is observed:

$$q(f) = p(f|D) = p(D|f)p(f)/p(D) \qquad (6)$$

Expanding the above equation is known as greatest posteriori estimation (P.S.Jagadeesh Kumar, T.Nagarajan, R.Srinivasan, Mingmin Pan, 2018). The decision of the earlier dissemination has as a rule been a tradeoff between a practical appraisal of convictions and picking a parametric structure that rearranges logical counts. By and by, specific types of the earlier are favored due principally to their numerical tractability. All discourses above depend on the objective of finding a point gauge of the model. In the Bayesian methodology, usually useful to have a choice capacity that considers the vulnerability of the model itself. A Bayesian prescient classifier is definitely for this reason:

$$f_{Bayes}(x) \triangleq E_{q(f)}\,[f(x)] \qquad (7)$$

In other words, rather than utilizing one point-gauge of the model, it is thought about the whole back dissemination, along these lines settling on the arrangement choice less subject to the fluctuation of the model. The utilization of

Bayesian prescient classifiers clearly prompts an alternate learning objective; it is currently the back conveyance *q(f)* that can be evaluated instead of a specific *f*. Accordingly, the preparation objective interchanges toward becoming $R_{p(x,y)}(q(f))$. Like the prior discourse, this goal can be judged by means of exact hazard minimization with regularization. For instance, McAllester's PAC-Bayesian bound acclaims the accompanying preparing objective,

$$q^*(f) = arg_q \min(E_{q(f)}[R_{emp}(f)] + \lambda D(q(f) \| p(f)) \quad (8)$$

This originates a back appropriation that limits both the minimized exact hazard just as the uniqueness from the earlier dissemination of the model.

3 ASR and Machine Learning

From a practical view, ASR is the transformation procedure from the acoustic information arrangement of discourse into a word grouping (G.Zhou, J.H.L.Hansen, and J.F.Kaiser, 2001). From the specialized perspective on ML, this transformation procedure of ASR requires various sub-forms including the utilization of discrete time stamps, frequently called edges, to describe the discourse waveform information or acoustic highlights, and the utilization of clear cut marks to file the acoustic information grouping. The principal issues in ASR lie in the idea of such names and information. It is critical to obviously comprehend the one of a kind properties of ASR, as far as both information and yield names, as a focal inspiration to associate the ASR and ML look into territories and to value their cover. From the yield perspective, ASR yields sentences that comprise of a variable number of words. In this manner, at any rate on a fundamental level, the quantity of conceivable classes for the characterization is large to the point that it is for all intents and purposes difficult to develop ML models for complete sentences without the utilization of structure.

From the data perspective, the acoustic information are besides an arrangement with a variable length, and usually, the length of information input is limitlessly not the same as that of mark yield, offering ascend to the uncommon issue of division or arrangement that the "static" grouping issues in ML don't experience. Consolidating the information and yield perspectives, the principal issue is expressed as an organized arrangement order task, where a succession of acoustic information is utilized to construe a grouping of the semantic units, for example, words. It is important that the grouping structure in the yield of ASR is commonly more perplexing than the vast majority of order issues in ML where the yield is a fixed, limited arrangement of classes (N.Fragopanagos and J.G.Taylor, 2005). Further, when sub-word units and setting reliance are acquainted with develop organized models for ASR, significantly more noteworthy intricacy can emerge than the clear word procedure yield in ASR. All the more intriguing and one of a kind issue in ASR, be that as it may, is on the input side, specifically, the variable-length acoustic grouping. The characteristic normal for discourse as the acoustic contribution to ML calculations makes it an occasionally more troublesome article for the examination than other. In that capacity, in the commonplace ML writing, there has ordinarily been less accentuation on discourse and related transient designs than on different flags and examples. The special normal for discourse lies basically in its transient measurement; specifically, in the tremendous fluctuation of discourse related with the flexibility of this worldly measurement (P.S.Jagadeesh Kumar, Yanmin Yuan, Yang Yung, Mingmin Pan, Wenli Hu, 2018). As a concern, regardless of whether two yield word successions are indistinguishable, the information discourse information regularly have particular lengths; e.g., distinctive information tests from a similar sentence more often than not contain diverse information dimensionality relying upon how the discourse sounds are delivered. Further, the discriminative signs among isolated discourse classes are often appropriated over a sensibly long worldly range, which regularly crosses neighboring discourse units. Other unique parts of discourse include class-subordinate acoustic signals (R.W.Picard, 1995). These signals are regularly communicated over various time traverses that would profit by various lengths of examination windows in discourse investigation and highlight extraction. At long last, recognized from other grouping issues generally examined in ML, the ASR issue is a unique class of organized example acknowledgment where the perceived examples, are implanted in the general transient arrangement design. Customary way of thinking places that discourse is a one-dimensional worldly flag as opposed to picture and video as higher dimensional signs. This view is generalized and does not catch the embodiment and challenges of the ASR issue. Discourse is best seen as a two-dimensional flag, where the spatial and fleeting measurements have unfathomably unique qualities, as opposed to pictures where the two spatial measurements will in general have comparative properties. The spatial measurement in discourse is related with the recurrence propagation and related changes, catching various variation types including essentially those emerging from situations, speakers, emphasize, talking style and rate. The last sort instigates relations amid spatial and worldly measurements, and the earth factors include amplifier qualities, discourse transmission, surrounding commotion, and room resonation.

The fleeting measurement in discourse, and specifically its relationship with the spatial or recurrence area properties of discourse, establishes one of the remarkable difficulties for ASR (Y.Bengio, 2009). A portion of the progressed generative models interrelated with the generative learning worldview of ML, where Bayesian practices are utilized to give fleeting limitations as earlier information about the human discourse age process. Utilizing the idea of the misfortune work just as the choice capacity, the significant ML ideal models are separated into generative and discriminative learning classifications. Contingent upon what sort of preparing information is accessible for learning, on the other hand ML ideal representations are systematized into directed, semi-managed, unsupervised, and dynamic learning classes (P.S.Jagadeesh Kumar, Wenli Hu, Yang Yung, 2017). At the point when dissimilarity among source and target appropriations emerges, a more typical condition in ASR than numerous different zones of ML applications, the ML ideal models are arranged into single-task; perform various tasks, and versatile learning. At last, utilizing the characteristic of information portrayal has scanty learning and deep learning ideal models, both later advancements in Machine Learning and Automatic Speech Recognition.

4 Generative and Discriminative Learning

Generative learning and discriminative learning are the two most pervasive, unfairly combined ML standards created and conveyed in ASR. There are two key factors that recognize generative gaining from discriminative learning: the nature of the model and the misfortune work. Quickly, generative learning comprises of utilizing a generative model, and embracing a preparation target work dependent on the joint probability misfortune characterized on the generative model (M.Caza-Szoka, D.Massicotte, and F.Nougarou, 2015). Discriminative learning, then again, requires either utilizing a discriminative model, or applying a discriminative preparing target capacity to a generative model. While generally there has been a solid relationship between a model and the misfortune work picked to prepare the model, there has been no fundamental matching of these two parts in the writing. This section will offer a decoupled perspective on the models and misfortune works ordinarily utilized in ASR to illustrate the natural relationship and difference between the ideal models of generative versus discriminative learning. Likewise, will exhibit the half and half learning worldview built utilizing blended generative and discriminative learning. In ASR, the most well-known generative learning approach depends on Gaussian-Mixture-Model based Hidden Markov models, or GMM-HMM. A GMM-HMM is parameterized by $\lambda = (\pi, A, B)$. π is a vector of state earlier probabilities; $A = (a_{i,j})$ is a state progress likelihood grid; and $b = \{b_1, \ldots, b_n\}$ is where speaks to the Gaussian blend model of state j. The state is commonly connected with a sub-fragment of a telephone in discourse. One significant development in ASR is the presentation of setting subordinate states, roused by the longing to lessen yield fluctuation related with each express, a typical system for "point by point" generative demonstrating (Y.Cho and L.K.Saul, 2009). A result of utilizing setting reliance is tremendous developments of the HMM state space, which, luckily, can be constrained by regularization techniques, for example, state tying. The presentation of the HMM and the related factual techniques to ASR in mid 1970s, can be respected the most huge change in outlook in the field, as examined in. One noteworthy purpose behind this early achievement was expected to the exceedingly productive MLE strategy imagined around ten years sooner. This MLE technique, regularly called the Baum-Welch calculation, had been the essential method for preparing the HMM-based ASR frameworks until 2002, is as yet one noteworthy advance in preparing these frameworks these days. It is fascinating to take note of that the Baum-Welch calculation fills in as one noteworthy inspiring case for the later improvement of the more broad Expectation-Maximization (EM) calculation. The objective of MLE is to limit the experimental hazard about the joint probability misfortune (reached out to consecutive information), i.e.

$$R_{emp}(f) = -\sum_i \ln p\left(x^{(i)}, y^{(i)};\, \pi, A, B\right) \qquad (9)$$

Generally as a succession highlight vectors extricated at edge level; communicates to a grouping of semantic units. In extensive vocabulary ASR frameworks, it is regularly the situation that word-level names are given, while state-level names are inactive (G.E.Hinton, R.Salakhutdinov, 2006). In addition, in preparing HMM-based ASR frameworks, parameter tying is often utilized as a sort of regularization. For instance, relative acoustic conditions of the triphones can have the equivalent Gaussian blend model. For this condition, the term *C(f)* is shifted by

$$C(f) = \prod_{(m,n)\epsilon T} \delta(b_m = b_n) \qquad (10)$$

The utilization of the generative model of HMMs, including the most mainstream Gaussian-blend HMM, for speaking to the unique discourse design and the utilization of MLE for preparing the tied HMM parameters establish one most conspicuous and fruitful case of generative learning in ASR. This achievement was immovably settled by the ASR people group, and has been broadly spread to the ML and related networks; truth be told, HMM has turned into a standard instrument in ASR as well as in ML and their related fields, for example, bioinformatics and characteristic language handling. For some ML just as ASR specialists, the achievement of HMM in ASR is somewhat amazing because of the outstanding shortcomings of the HMM (P.Dayan, 2009). Another reasonable achievement of the generative learning worldview in ASR is the utilization of GMM-HMM as earlier "information" inside the Bayesian structure for condition vigorous ASR. At the point when the discourse motion, to be alleged, is blended with commotion or another non-expected speaker, the perception is a mix of the flag of intrigue and impedance of no intrigue, both obscure. Without earlier data, the recovery of the discourse of intrigue and its acknowledgment would be not well described and subject to net blunders. Exploiting generative models of Gaussian-blend HMM (additionally filling the double need of recognizer), or frequently a less complex Gaussian blend or even a solitary Gaussian, as Bayesian earlier for "clean" discourse defeats the badly presented issue. Further, the generative method permits probabilistic development of the model for the relationship among the speech perception, clean discourse, and impedance, which is normally nonlinear when the log-area highlights are utilized. A lot of generative learning approaches in ASR following this rationality are fluidly called "parallel model mix", vector Taylor procedure. Strikingly, the exhaustive use of such a generative learning worldview for single-channel multitasked discourse acknowledgment is accounted for and looked into where the creators apply effectively various entrenched ML strategies including loopy conviction engendering and organized mean-field estimate. Utilizing this generative learning plan, ASR precision with uproarious meddling speakers is appeared to surpass human execution.

At the point when connected to ASR, there are instant approaches which utilize most extreme entropy Markov models, restrictive irregular fields, hidden Conditional Random Fields (hCRFs), increased CRFs, segmental CRFs, and profound organized CRFs. The utilization of neural systems as MLP (ordinarily with one shrouded layer) with the softmax nonlinear capacity at the last layer was well known in 1990's. Since the yield of the MLP can be deciphered as the restrictive likelihood, when the yield is bolstered into a HMM, a prodigious discriminative grouping model, or cross breed MLP-HMM, can be prepared. The utilization of this kind of discriminative model for ASR has been reported. Due for the most part to the trouble in learning MLPs, this line of research has been changed to another course where the MLP just creates a subset of "include vectors" in mix with the conventional highlights for use in the generative HMM (P.S.Jagadeesh Kumar, 2018). Recently, the trouble related with learning MLPs has been effectively tended. Every model is instances of the probabilistic discriminative models conversed as restrictive prospects of discourse classes given the acoustic highlights as the input. The second school of discriminative models centers on choice parameters rather than class-restrictive probabilities. Practically equivalent to MLP-HMMs, SVM-HMMs have been created to give progressively precise state/telephone grouping scores, with intriguing outcomes. Ongoing work has endeavored SVMs and have gotten huge execution gains in commotion strength ASR.

5 Semi-Supervised and Active Learning

Supervised learning expects that all preparation tests are named, while unsupervised learning accepts none. Semi-supervised learning, as the name recommends, expect that both named and unlabeled preparing tests are accessible. Supervised, unsupervised and semi-supervised learning are normally alluded to under the aloof getting the hang of setting, where named preparing tests are produced unevenly as per an obscure likelihood conveyance. Interestingly, dynamic learning is where the learner can astutely pick which tests to appellation (L.C.Resende, L.A.F.Manso, W.D.Dutra, A.M.L.da Silva, 2012). In this segment the focus is for the most part on semi-supervised and dynamic learning ideal models. This is on the grounds that regulated learning is sensibly known and unsupervised learning does not straightforwardly go for anticipating yields from data sources. Firstly, call attention to that the standard depiction of semi-managed learning talked about above in the ML writing has been utilized freely in the ASR writing, and frequently been alluded to as unsupervised learning or unsupervised preparing. This disarray is brought about by the way that while there are both translated/marked and un-interpreted arrangements of preparing information, the last is fundamentally more noteworthy in the sum than the previous. In fact, the requirement for semi-directed learning in ASR is self-evident. Cutting edge execution in expansive vocabulary ASR frameworks as a rule requires a huge number of long stretches of physically clarified discourse and a great many expressions of content. The manual translation is regularly excessively costly or unrealistic. Luckily, one can depend upon the supposition that any area which requires ASR innovation will have a huge number of long stretches of sound accessible. Unsupervised acoustics model preparing fabricates introductory models from little measures of deciphered acoustic information and after that utilization of them to interpret a lot bigger measures of un-translated information (Y.Bengio, 2009). One at that point prepares new models utilizing part or these programmed transcripts as the name. This definitely diminishes the naming prerequisites for ASR in the meager areas. The above preparing worldview cascades into oneself preparing class of semi-supervised learning. Agent work incorporates, where an ASR prepared on a little interpreted set is utilized to produce interpretations for bigger amounts of un-translated information first. The perceived translations are chosen at that point dependent on certainty measures. The chose translations are treated as the right ones and are utilized to prepare the last recognizer.

Explicit strategies incorporate steady preparing where the high certainty (as decided with a limit) expressions are joined with interpreted articulations to retrain or to adjust the recognizer. At that point the retrained recognizer is utilized to interpret the following group of expressions. Regularly, summed up desire amplification is utilized where all expressions are utilized yet with various loads controlled by the certainty measure (C.Fredouille, G.Pouchoulin, J.-F.Bonastre, M.Azzarello, A. Giovanni, and A.Ghio, 2005). This methodology fits into the general structure has additionally been connected to consolidating discriminative preparing with semi-managed learning. While direct, it has been appeared such certainty based self-preparing approaches are related with the limitation of strengthening what the present model definitely knows; now and then notwithstanding fortifying the mistakes. Dissimilarity is much of the time seen when the execution of the present model is generally poor. Like the goal of, in crafted by the worldwide entropy considered over the whole preparing informational index is utilized as the reason for allocating marks in the un-deciphered part of the preparation expressions for semi-directed learning. This methodology varies from the past ones by settling on the choice dependent on the worldwide dataset rather than individual articulations as it were. All the more explicitly, the created calculation centers around the improvement to the general framework execution by contemplating the certainty of every articulation as well as the recurrence of comparable and opposing examples in the un-deciphered set while deciding the correct expression interpretation pair to be incorporated into the semi-directed preparing set. The calculation assesses the normal entropy decrease which the articulation translation pair may cause on the full un-deciphered dataset. Other ASR work in semi-managed learning influences earlier information, e.g., shut subtitles, which are considered as low-quality or loud marks, as limits in generally standard self-preparing. One specific limitation abused is to regulate the closed inscriptions to perceived translations and to choose just fragments that concur. This methodology is called daintily managed. On the other hand, acknowledgment has been completed by using a language model which is prepared on the shut inscriptions. One might bring up that numerous sustainable semi-managed learning controls created in ML but still can't seem to be investigated in ASR, and this is one zone anticipating developing commitments from the ML group.

Active learning is a comparative setting to semi-directed learning. The objective of dynamic adapting, be that as it may, is to question the most instructive arrangement of contributions to be marked, planning to improve grouping execution with the base number of inquiries. That is, in dynamic learning, the learner may assume a functioning job in choosing the informational index as opposed to it is inactively given. The key thought behind dynamic learning is that a ML calculation can accomplish more noteworthy implementation, e.g., higher grouping exactness, with less preparing names on the off chance that it is permitted to pick the subset of information that has names (J.T.Senders, 2018). A functioning learner may present investigations, as a rule as unlabeled evidence examples to be marked (frequently by a human). Hence, it is in some cases called question learning. Dynamic learning is well-spurred in many advanced ML issues, where unlabeled information might be plenteous or effectively gotten, yet marks are troublesome, tedious, or costly to get. This is the context for discourse response. Extensively, dynamic learning comes in two structures: clump dynamic realizing, where a subset of information is picked, from the earlier in a cluster to be named. The names of the examples in the clump picked to be marked may not, under this methodology, impact different occurrences to be chosen since all cases are picked without a moment's delay. In online dynamic learning, then again, occurrences are picked one-by-one, and the genuine marks of all recently named cases might be utilized to choose different examples to be named. Thus, online dynamic learning is once in a while considered all the more dominant.

6 Adaptive and Multi-task Learning

Adaptive learning, or transfer learning with "information exchange", is another machine learning worldview that underlines creating a classifier that sums up crosswise over propagations, spaces, or assignments. Exchange learning is increasing developing significance in ML as of late yet is all in all less well-known to the ASR people group than other learning standards examined up until this point. In fact, various profoundly fruitful adjustment strategies created in ASR are intended to illuminate a standout amongst the most unmistakable issues that move learning specialists in ML endeavor to address; confuse among preparing and test conditions. Be that as it may, the extent of move learning in ML is more extensive than this, and it likewise includes various plans natural to ASR specialists, for example, broad media ASR, multi-lingual and cross-lingual ASR, speech learning for word acknowledgment, and identification based ASR. The sort out such differing ASR procedures into a bound together classification plot under the extremely wide move was learning worldview, which would somehow or another be seen as segregated ASR applications. Likewise, the standard ML documentations to portray all ASR themes are utilized (P.S.Jagadeesh Kumar, 2017). There is huge ML writing on exchange learning. To arrange the overview with surveys to existing ASR uses, four-route classification of major exchange learning strategies is made, utilizing the accompanying two tomahawks. The principal pivot is the way in which information is exchanged. Versatile learning is one type of move learning in which information move is done in a successive way, usually from a source errand to an objective assignment. Interestingly, perform various tasks learning is worried about learning different errands at the same time. Adaptive learning can be symmetrically sorted utilizing the second hub about whether the info/yield space of the objective errand is not the same as that of the source task. It is called homogeneous if the source and target task have a similar yield space, and is heterogeneous. Note that both adaptive learning and multiple tasks learning can be either homogeneous or heterogeneous.

The terms heterogeneous exchange and perform multiple tasks learning are frequently utilized exchangeably in the ML writing, as perform various tasks adapting for the most part includes heterogeneous sources of info or yields, and the data conversation can go the two headings between assignments. One most fascinating use of heterogeneous exchange and perform various tasks learning is Multi-task Learning, just as acknowledgment and amalgamation of different wellsprings of methodology data, for example, video and picture. In the ongoing investigation, a case of heterogeneous multi-solicit taking in engineering from is created utilizing further developed various leveled models and deep learning methods (A.Mukherjee, A.Pal, P.Misra, 2012). This deep learning model is then connected to various undertakings including discourse acknowledgment, where the sound information of discourse (as spectrogram) and video information are intertwined to become familiar with the mutual portrayal of both discourse and video in the mid layers of a speech engineering. This performs multiple tasks profound design broadens the prior deep models bent for single-task deep learning engineering for picture pixels and for discourse spectrograms. The fundamental outcomes demonstrate that both video and discourse acknowledgment undertakings are improved with perform various tasks learning dependent on the profound designs empowering shared discourse and video portrayals. Another effective case of heterogeneous exchange and perform various tasks learning in ASR is multi-lingual or cross-lingual discourse response, where discourse response for various dialects is considered as various undertakings (P.R.Krugman, M.Obstfeld, and M.J.Melitz, 2012). Diverse methodologies have been taken to assault this somewhat testing acoustic demonstrating issue for ASR, where the trouble lies in low assets in either information or interpretations or both because of monetary contemplations in creating ASR for all dialects of the world. Cross-language information sharing and information weighing are normal and valuable methodologies. Another fruitful methodology is to outline units crosswise over dialects either through learning based or information driven strategies. At long last, considering telephone acknowledgment and word response as various errands, e.g., telephone acknowledgment results are utilized not for delivering content yields but rather for language-type recognizable proof or for spoken archive recovery, at that point the utilization of elocution lexicon in practically all ASR frameworks to connect telephones to words can include another incredible case of heterogeneous exchange. Further developed systems in ASR have pushed this heading further by supporting the utilization of even better units of discourse than telephones to connect the crude acoustic data of discourse to semantic substance of discourse by means of a chain of command of phonetic structure (P.Domingos, 2012). These nuclear discourse units incorporate "discourse traits" in the discovery based and information rich demonstrating structure, and covering articulatory highlights in the system that empowers the misuse of articulatory restraints and discourse co-articulatory modules for familiar discourse acknowledgment. At the point when the articulatory data amid discourse can be recuperated amid discourse acknowledgment utilizing articulatory based recognizers, such data can be applicably connected to an alternate task of vocalization preparing.

7 Deep Learning and ASR Techniques

Deep learning alludes to a class of ML systems, where numerous layers of data preparing stages in various leveled structures are misused for unsupervised component learning and for example characterization. It is in the crossing points among the exploration regions of neural system, graphical demonstrating, advancement, design acknowledgment, and flag preparing (P.S.Jagadeesh Kumar, Yang Yung, Wenli Hu, 2017). Two major explanations behind the prevalence of profound adapting today are the essentially brought down expense of figuring equipment and the radically expanded chip handling capacities. Since 2006, specialists have exhibited the achievement of deep learning in assorted uses of computer vision, phonetic acknowledgment, voice seek, unconstrained discourse acknowledgment, discourse and picture include coding, semantic articulation description, hand-composing acknowledgment, sound handling, data recovery, and mechanical autonomy. As depicted before, deep learning alludes to a somewhat wide class of ML systems and structures, with the sign of utilizing numerous layers of non-direct data preparing stages that are various leveled in nature. Contingent upon how the structures and systems are proposed for use, e.g., amalgamation/age or acknowledgment/arrangement, one can order the vast majority of the work around there into three kinds outlined beneath (P.S.Jagadeesh Kumar, S.Meenakshi Sundaram, 2015). The main sort comprises of generative profound models, which are expected to describe the high-request relationship properties of the information or joint factual disseminations of the unmistakable information and their related classes. Utilization of Bayes standard can transform this sort of engineering into a discriminative one. Instances of this sort are different types of deep auto-encoders, deep Boltzmann machine, and its expansion to the figured higher-request Boltzmann machine in its base layer. Different types of generative models of shrouded discourse elements, the profound powerful Bayesian system model, likewise have a place with this sort of generative profound models. The second sort of profound designs are discriminative in nature, which are planned to give discriminative capacity to design characterization and to do as such by portraying the back circulations of class names molded on the noticeable information (C.Chelba, T.J.Hazen, and M.Saraclar, 2008). Models incorporate profound organized CRF, pair MLP engineering profound raised or stacking system and its tensor form and recognition based ASR design. In the third sort, or half breed profound designs, the objective is separation however this is helped with the results of generative structures. In the current mixture structures distributed in the writing, the generative part is for the most part misused to help with separation as the last objective of the half and half engineering. How and why generative demonstrating can help with discriminative can be studied from two outlooks:

1) The enhancement perspective where generative models can give magnificent introduction focuses in exceptionally nonlinear parameter estimation issues (The usually utilized term of "pre-preparing" in deep learning); and

2) The regularization point of view where generative models can adequately control the intricacy of the general model.

At the point when the generative profound design of DBN is liable to assist discriminative preparing, generally called "adjusting" in the writing, a balanced engineering of deep neural network (DNN, which is in some cases also called DBN or deep MLP). In a DNN, the loads of the system are "pre-prepared" from DBN rather than the standard irregular introduction. The astonishing attainment of this mixture generative-discriminative profound design as DNN in substantial vocabulary ASR, before long confirmed by a progression of new and greater ASR errands did vivaciously by various major ASR labs around the world. Another run of the mill case of the half breed profound design was created (F.Biadsy, 2011). This is a crossover of DNN with a superficial discriminative design of Conditional Random Fields (CRF). At this time, the general design of DNN-CRF is gotten the hang of utilizing the discriminative rule of sentence-level restrictive likelihood of names given the info statistics grouping. It tends to be appeared such DNN-CRF is identical to a half breed profound design of DNN and HMM, whose parameters are found out together utilizing the full-grouping greatest common data between the whole name arrangement and the information succession. This engineering is all the more as of late stretched out to have successive associations or transient reliance in the concealed layers of DBN, notwithstanding the yield layer. Displaying organized discourse elements and exploiting the basic fleeting properties of discourse are vital to high exactness ASR. However the DBN-DNN approach, while achieving sensational blunder decrease, has utilized such organized elements. Rather, it just recognizes the contribution of a long window of discourse includes as its acoustic setting and yields an expansive number of setting subordinate sub-telephone units, utilizing many concealed layers one over another with huge loads. The insufficiency in worldly parts of the DBN-DNN approach has been perceived and quite a bit of ebb and flow look into has focused on intermittent neural system utilizing the equivalent huge weight process (R.Schapire, 2008). It isn't clear such a beast power approach can enough catch the fundamental organized unique properties of discourse, however it is obviously better than the prior utilization, fixed-sized windows in DBN-DNN. The most effective technique to assimilate the thoughtfulness of generative demonstrating of discourse elements, into the discriminative deep structures investigated vivaciously by both ML and ASR people group as of late is a creative research bearing. Dynamic research is presently continuous by a developing number of congregations, both scholastic and modern, in concerning profound figuring out how to ASR. New and progressively successful deep structures and related learning calculations have been accounted for in each major ASR-related and ML-related gathering since 2015. This pattern is required to proceed in coming years.

As of late, another dynamic region of ASR explore that is firmly identified with ML has been the utilization of meager portrayal. This alludes to a lot of systems used to remake an organized flag from a set number of preparing precedents, an issue which emerges in numerous ML applications where reproduction identifies with adaptively finding a word reference which best speaks to the flag on a for each example premise. The word reference can either incorporate irregular projections, as is commonly proficient for flag remaking, or incorporate real preparing tests from the information, as investigated additionally in numerous ML applications. Like deep learning, scanty portrayal is another rising and quickly developing zone with obligations in an assortment of flag preparing and ML meetings, incorporating ASR as of late. The ongoing audit uses inadequate portrayal to ASR, featuring the importance to and commitments from ML (T.Bocklet, A.Maier, J. G.Bauer, F.Burkhardt, E.Nöth, 2008). In model based inadequate portrayals are deliberately examined to delineate highlights into the direct range of preparing precedents. They share the equivalent "nonparametric" ML standard as the closest neighbor approach investigated and the SVM technique in rightfully using data about individual preparing precedents. In particular, given a lot of acoustic-highlight groupings from the preparation set that fill in as a word reference, the test information is spoken to as a straight blend of these preparation precedents by taking care of a least square relapse issue obliged by meager condition on the weight preparation. The utilization of such requirements is ordinary of regularization strategies, which are key in ML. The inadequate highlights acquired from the meager loads and lexicons are then used to outline test once again into the straight range of preparing precedents in the word reference. The outcomes demonstrate that the casing level discourse order exactness utilizing scanty portrayals surpasses that of Gaussian blend model. Furthermore, inadequate portrayals not just draw test comprises nearer to preparing; they moreover draw the highlights nearer to the right class. Such meager portrayals are utilized as extra highlights to the current fantastic highlights and mistake rate decrease is accounted for both telephone acknowledgment and huge vocabulary constant discourse acknowledgment errands with point by point test conditions. Inadequate portrayal has close connects to central ML ideas of regularization and unsupervised element learning, and furthermore has a profound root in neuroscience (T.Vogt, E.Andro, 2006). In any case, its applications to ASR are very later and their prosperity, contrasted and deep learning, is progressively constrained in degree and size, regardless of the notable achievement of inadequate coding and compressive detecting in ML and flag/picture preparing with a moderately long history. One plausible constraining component is that the hidden structure of discourse highlights is less inclined to sparsification and pressure than the picture partner. In any case, the underlying promising ASR results as checked on above ought to empower more work toward this path. It is conceivable that diverse sorts of crude discourse highlights from what have been tested will have more notable potential and adequacy for inadequate portrayals. For instance, discourse waveforms are clearly not a characteristic possibility for scanty portrayal but rather the leftover flags after direct expectation would be. Further, meager condition may not really be abused for portrayal purposes just in the unsupervised getting the hang of setting. Similarly as the accomplishment of profound taking in originates from half and half amid unsupervised generative learning (pre-preparing) besides directed discriminative adapting (tweaking), meager condition can be misused along these lines. The ongoing work defines parameter meager condition as delicate regularization and curved obliged advancement issues in a DNN framework. Rather than putting meager condition requirement in the DNN's shrouded hubs for highlight portrayals, inadequacy is abused for decreasing non-zero DNN loads (I.M.A.Shahin, 2013). The exploratory outcomes on an extensive scale ASR task show not just the DNN model size is decreased by 66% to 88%, the blunder rate is marginally diminished by 0.2–0.3%. It is a prolific research bearing to misuse inadequacy in different ways for ASR, and the very active profound coding plans created by ML and computer vision specialists presently can't seem to enter ASR.

8 Results and Implication

Various machine learning techniques for automatic speech recognition were clearly observed and the one relevant for automatic speech recognition based robotic surgery were mainstreamed. One good thing about robotic surgery is that the environment and other atmospheric conditions are least concerned as the surgery is conducted in calm and pleasant environment. The authors neglected the negligible factors like atmospheric turbulence and environmental noise as the surgery is conducted in indoor conditions. It was observed that most of the traditional machine learning algorithms is not suitable for automatic speech recognition except few namely Hidden Markov Models, Conditional Random Fields, Gaussian model, etc. It was found that those few traditional machine learning techniques were not able to provide the required accuracy and efficiency in the case of ASR. Deep learning technique was a windfall for speech recognition arena irrespective of few considerable setbacks. Unlike real-time speech recognitions like phonic dialog and speaker appreciation, the speech recognition aspects in medical conditions are different. The robot has to be trained with respect to both local language vocabulary and medical lexicon. But to relax, most of the robotic surgeries are applied only for selected practice unlike human experts. Robotic surgery is a practice to realize surgery by means of small tools devoted to a robotic arm. The surgeon controls the robotic arm with a computer. The new research direction stated in this paper is to control the robotic arm through speech processing unlike computer programming. For most of the time as of now, the robotic arm movements have been controlled by the surgeon hand movement. In the last few years extensive research had been conducted by Stanford Medical Research Centre in this direction. Table 1, Fig. 1 and Fig.2 demonstrates deep learning based automatic speech recognition of robotic brain surgery for cancer. The results clearly demonstrates the supremacy of deep learning but to the revelation, combination of deep learning and traditional machine learning methods provided even more better accuracy. Hybrid deep learning method like Deep Neural Network-Hidden Markov Models (DNN-HMM) provided 82.5% word error rate. Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system.

9 Conclusion

In this paper, a lot of prestigious ML models that are bothered in the milieu of ASR mastery and applications are built up just as ML is powerfully inherent inside ASR innovation, and the other way around. ASR can be viewed as just as an instance of Machine Learning issue, similarly as a couple of germaneness of ML like computer vision, bioinformatics, and robotics as well. ASR is chiefly useful ML application since it has enormously colossal preparing and testing designs, it is computationally critical, it has a supreme sequential plan in the information, it is additionally a case of ML with organized yield, and perchance most altogether, it has a gigantic culture of assessors who are vivaciously advancing in the hidden inclination. ASR has been the premise to a ton of noteworthy musings in ML. Unquestionably; the key reason is that these two gatherings can and ought to relate as often as possible with one another. The trust is that the authentic and mutually ideal gatherings impact the social orders that had on one another will proceed, maybe at a much progressively productive pace. It is trusted that this paper will without a doubt encourage such correspondence and progression. To this end, the key ML idea of organized arrangement as a basic issue in ASR; as for both the emblematic succession as the ASR classifier's yield and the ceaseless esteemed vector include grouping as the ASR classifier's info are explained. In displaying every one of the ML ideal models, the most important ML ideas to ASR are featured, and the sort of ML approaches that are compelling in managing the exceptional challenges of ASR including profound/unique structure in human discourse and solid fluctuation in the perceptions is underlined. Exceptional consideration is paid to talk about and investigate the significant ML ideal models and results that have been affirmed by ASR tests. The primary models talked about in this article incorporate generative learning and discriminative learning, versatile and Bayesian learning for condition vigorous and speaker-powerful ASR, and half and partially regulated/unsupervised learning or crossover generative/discriminative learning as publicized in the later "Deep learning" conspire. ASR revolution is quick changing as of late, rather pushed by various developing applications in portable processing, and AI-like individual associate innovation. So is the mixture of ML strategies into ASR. A complete analysis on the subject of this nature inevitably contains inclination as it recommends significant research issues and future headings where the ML standards would offer the opportunity to goad next floods of ASR evolution. Later on, increasingly incorporated ML standards to be conveniently connected to ASR as exemplified by the two developing ML plans are normal. New ML systems that utilize extensive supply of preparing information with wide decent variety and expansive scale improvement to affect ASR, where dynamic learning, semi-directed learning, and even unsupervised adapting not well assumed more lavish jobs than previously and at present are overviewed. In addition, viable inquiry and taking advantage of deep learning, numerous leveled structures related to spatially invariant and transitory powerful properties of discourse is simply preparatory. The ongoing reestablished enthusiasm for repetitive neural networks with deep learning, different dimension portrayals from both ASR and ML group using more dominant streamlining strategies than in the past is a case of the examination moving towards this direction. To procure full organic product by such an undertaking will require incorporated ML procedures inside and potentially past the ideal models.

To conclude, automatic speech recognition based robotic surgery with no doubt is going to rule surgical field in the decades to come. Though, hybrid deep learning based speech recognition provided considerable word accuracy, the authors' future research perspective is to identify a dedicated algorithm for robust automatic speech recognition based robotic surgery to fast-track required accuracy and economically affordable technology.

References

M.Dhanalakshmi, T.Nagarajan, P.Vijayalakshmi: Speech-Input Speech-Output Communication for Dysarthric Speakers Using HMM-Based Speech Recognition and Adaptive Synthesis System. *CSSP* 37(2): pp.674-703 (2018).

D.Reynolds.: An overview of automatic speaker recognition technology. In Proc. *IEEE Int. C. Acoust., Speech Signal Process*., vol. 4, pp. 4072–4075, (2002).

S.Furui.; Speaker-dependent-feature extraction, recognition and processing techniques. *Speech Commun*. vol. 10, nos. 5–6, pp. 505–520, (1991).

G.Zhou, J.H.L.Hansen, and J.F.Kaiser.: Nonlinear feature based classification of speech under stress. *IEEE Trans. Speech Audio Process*. vol.9, no.3, pp. 201–216, (2001).

P.S.Jagadeesh Kumar, Yanmin Yuan, Yang Yung, Mingmin Pan, Wenli Hu.: Robotic Simulation of Human Brain Using Convolutional Deep Belief Networks. *International Journal of Intelligent Machines and Robotics*, 1 (2), pp. 180-191, (2018).

C.Fredouille, G.Pouchoulin, J.-F.Bonastre, M.Azzarello, A. Giovanni, and A.Ghio. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). In Proc. *Eur. Conf. Speech Commun. Technol. (Eurospeech)*, pp. 149–152, (2005).

R.W.Picard.: Affective Computing. MIT Press, Cambridge, MA, USA, *Tech. Rep*. 321, pp. 1–16, (1995).

P.S.Jagadeesh Kumar, Yang Yung, Wenli Hu.: Tensor Deep Stacking Neural Networks and Bilinear Mapping Based Speech Emotion Classification Using Facial Electromyography. *International Conference on Speech Recognition and Language Processing*, Amsterdam, Netherlands, December 04-05, (2017).

E.Douglas-Cowie, N.Campbell, R.Cowie, and P. Roach.: Emotional speech: Towards a new generation of databases. *Speech Commun*., vol. 40, nos. 1–2, pp. 33–60, (2003).

E.Nöth, S.Harbeck, and H.Niemann.: Multilingual speech recognition in Computational Models of Speech Pattern Processing, *NATO ASI Series*, vol. 169. Springer, 1999, pp. 362–374, (1999).

T.Schultz, A.Waibel.: Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Commun*. vol. 35, no.2, pp. 31–51, (2001).

V.A.Petrushin.: Emotion recognition in speech signal: Experimental study, development, and application. In Proc. *6th Int. C. Spoken Lang. Process*. (2000).

N.Fragopanagos and J.G.Taylor.: Emotion recognition in human–computer interaction. *Neural Netw*., vol. 18, no. 4, pp. 389–405, (2005).

P.S.Jagadeesh Kumar, T.Nagarajan, R.Srinivasan, Mingmin Pan.: Smart Deep Learning Grounded Automatic Speech Recognition meant for Brain Arteriovenous Malformation. *IERI Int. Conference on Medical Physics, Medical Engineering and Informatics* (ICMMI2018), September 7-9, (2018).

C.Chelba, T.J.Hazen, and M.Saraclar.: Retrieval and browsing of spoken content. *IEEE Signal Process. Mag*., vol. 25, no. 3, pp. 39–49, May (2008).

F.Biadsy.: Automatic dialect and accent recognition and its application to speech recognition. Ph.D. thesis, Graduate School Arts Sci., *Columbia Univ*., New York City, NY, USA, 2011, pp. 1–171, (2011).

T.Bocklet, A.Maier, J. G.Bauer, F.Burkhardt, E.Nöth.: Age and gender recognition for telephone applications based on GMM supervectors and support vector machines. In Proc. *IEEE Int. Conf. Acoust. Speech Signal Process*, Apr. 2008, pp. 1605–1608, (2008).

T.Vogt, E.Andro.: Sanitizing automatic emotion recognition from speech via gender differentiation. In Proc. *Lang. Resour. Eval. Conf*., Jan. 2006, pp. 1123–1126, (2006).

I.M.A.Shahin.: Gender-dependent emotion recognition built on HMMs and SPHMMs. *Int. J. Speech Technol*., vol. 16, no. 2, pp. 133–141, (2013).

R.Schapire.: Theoretical Machine Learning. Princeton, NJ, USA: *Princeton Univ*., 2008, pp. 1–6, (2008).

P.Domingos.: A few useful things to know about machine learning. *Commun. ACM*, vol. 55, no. 10, pp. 78–87, (2012).

A.Mukherjee, A.Pal, P.Misra.: Data analytics in ubiquitous sensor based health information systems. In Proc. *6th Int. C Next Generation Mobile Appl. Services Technol.* (NGMAST), 2012, pp. 193–198, (2012).

P.S.Jagadeesh Kumar.: Modern Machine Learning Conniptions for Automatic Speech Recognition. *International Conference on Systems, Signals, Speech and Image Processing* ICSSSI'17, Los Angeles, USA, 09-10 October, (2017).

J.T.Senders et al.: An introduction and overview of machine learning neurosurgical care. *Acta Neurochirurgica*, vol. 160, no. 1, pp. 29–38, (2018).

P.R.Krugman, M.Obstfeld, and M.J.Melitz.: International Economics: *Theory and Policy*. New York, NY, USA: Prentice-Hall, (2012).

L.C.Resende, L.A.F.Manso, W.D.Dutra, A.M.L.da Silva.: Support vector machine application in composite reliability assessment. In Proc. *18th Int. Conf. Intell. Syst. Appl. Power Syst.* (ISAP), Portugal, (2015).

R.Rojas.: Unsupervised learning and clustering algorithms. In *Neural Networks*. Springer, pp. 99–121, (1996).

P.S.Jagadeesh Kumar, Wenli Hu, Yang Yung.: Virtual Reality Based 3D Video Games and Speech-Lip Synchronization Superseding Algebraic Code Excited Linear Prediction. *Int. Journal of Computer and Information Sciences* 4(12), pp.303-321, Dec 2017 *International Conference on Games and Simulations*, Kuala Lumpur, Malaysia, (2017).

M.Caza-Szoka, D.Massicotte, and F.Nougarou.: Naive Bayesian learning for small training samples: Application on chronic low back pain diagnostic with sEMG sensors. In Proc. *IEEE Int. Instrum. Meas. Technol. Conf.* (I2MTC), May 2015, pp. 470–475, (2015).

P.S.Jagadeesh Kumar.: Deep Learning Based Cognitive Self-Supervised Robots in Medical Engineering. *6th World Convention on Robots, Autonomous Vehicles* *and Deep Learning*, Smart Robotics Congress 2018, September 10-11, (2018).

Y.Cho and L.K.Saul.: Kernel methods for deep learning. In Proc. *Adv. Neural Inf. Process. Syst.* (NIPS), vol. 22, 2009, pp. 342–350, (2009).

P.Dayan.: Unsupervised learning. In *The MIT Encyclopedia of the Cognitive Science*s. Cambridge, MA, USA: MIT Press, 2009, pp. 1–7, (2009).

Y.Bengio.: Learning deep architectures for AI Found. *Trends Mach. Learn*., vol.2, no.1, pp.1–127, (2009).

G.E.Hinton, R.Salakhutdinov.: Reducing the dimensionality of data with neural networks. *Science*, vol. 313, no. 5786, pp. 504–507, (2006).

P.S.Jagadeesh Kumar, S.Meenakshi Sundaram.: Adapted Optimal Neural Network Based Classifier Using Optical Character Recognition Engine for Tamil Language. *International Journal in Foundations of Computer Science & Technology*, Vol.5, Issue 4, (2015).

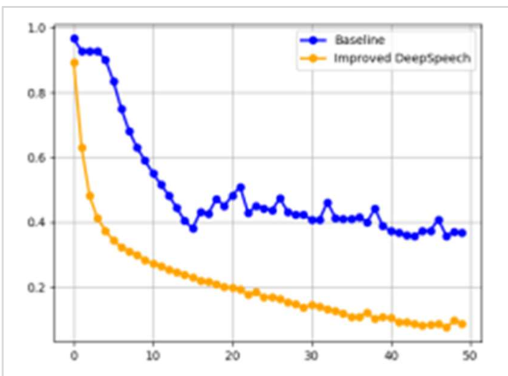

Fig. 1. Relative Gain (X-axis) Vs Word Error Rate (Y-axis) for baseline Automatic Speech Recognition and Deep Neural Network (DNN).

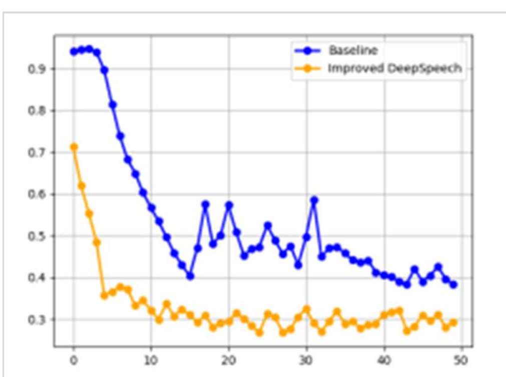

Fig. 2. Relative Gain (X-axis) Vs Word Error Rate (Y-axis) for baseline Automatic Speech Recognition and Deep Neural Network-Hidden Markov Models DNN-HMM.

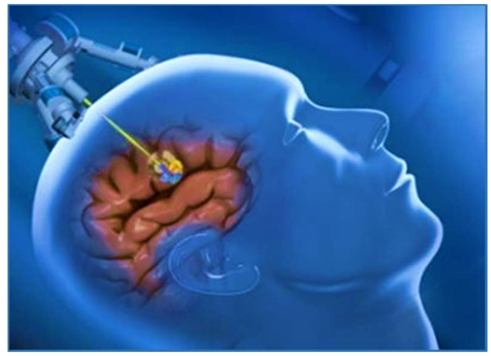

Fig. 3. Robotic surgery for treating brain cancer at Stanford Medical Research Centre utilizing speech recognition

Table 1. Deep learning based automatic speech recognition for robotic brain surgery (**Fig. 3**)

S.No.	Technique*	Method	Advantage	Disadvantage	Role in Automatic Speech Recognition (ASR)	Word Error Rate (WER) in percentage for Robotic Surgical Applications
1	**Generative Learning**	Naïve Bayes	Minimum training data	Loss of accuracy	Not Recommended	NA
		Bayesian networks	Predict according to the posterior	Erratic for Automatic applications	Not Recommended	NA
		Markov random fields (MRFs)	Compactly represent independence assumptions	Difficult to interpret	Not Recommended	NA
		Hidden Markov Models (HMMs)	Strong statistical foundation	Expensive in terms of memory and compute time	Recommended	60.8
2	**Discriminative Learning**	Logistic regression	Handles nonlinear effects	Less predictive performance	Not Recommended	NA
		Scalar Vector Machine	Non-parametric and operates locally	Lack of transparency	Not Recommended	NA
		Traditional neural networks	Ability to work with incomplete knowledge	Computationally expensive	Not Recommended	NA
		Nearest neighbor method	Ability to handle large training data	Sensitive to class-outliers	Not Recommended	NA
		Conditional Random Fields (CRFs)	Flexible feature selection	Complex re-training	Recommended	50.3
3	**Semi-Supervised Learning**	Transductive support vector machine (TSVM)	Unlabeled data is test data	Optimization becomes much harder to solve	Not Recommended	NA
		Graph-based method	Explicit mathematical formulation	Hard to construct graph in sparse spaces	Not Recommended	NA
		Self-training method	Higher confidence predictions	Frequent problem with noisy labels	Not Recommended	NA
		Co-training method	Training two classifiers which then teach each other	Require very specific setup	Not Recommended	NA

4	**Active Learning**	Support-vector machines (SVMs)	Ability to solve high dimensions and non-linearly separable problems	Several key parameters	Not Recommended	NA
		Minimum Marginal Hyperplane method	Optimal experimental design	Interactively queried	Not Recommended	NA
5	**Adaptive Learning**	Recursive Least Squares	Closed-loop solutions	In-stream analytics	Not Recommended	NA
6	**Multi-task Learning**	Clustered Multi-Task Learning	Sparse coding representation	Capture invariant properties	Not Recommended	NA
7	**Deep Learning**	Deep Neural Network (DNN)	Massive parallel computations	Extremely expensive to train due to complex data models	Recommended	78.6
		Convolutional Deep Neural Network (CDNN)	Feature learning	Needs a lot of training data	Recommended	71.8
		Bidirectional Recurrent Neural Network (BRNN)	Variable input data	Context of input data is not needed	Recommended	70.6
		Deep Belief Network (DBN)	Higher performance gain	Lack of flexibility	Recommended	69.5
		Deep Neural Network-Hidden Markov Models (DNN-HMM)	Avoids over-fitting issue of traditional neural networks	Needs a huge number of parameters to tune	Recommended	**82.5**
		Deep Neural Network-Conditional Random Fields (DNN-CRFs)	Lack of transparency in their thinking process	Model is over trained	Recommended	74.3
		Convolutional Deep Neural Network-Recurrent Neural Network (CDNN-RNN)	Progressive tendency to learn	Difficult to train	Recommended	72.5

*HMMs - Hidden Markov Models (Generative Learning)
*CRFs - Conditional Random Fields (Discriminative Learning)
*DNN - Deep Neural Network (Deep Learning)
*CDNN - Convolutional Deep Neural Network (Deep Learning)
*BRNN - Bidirectional Recurrent Neural Network (Deep Learning)
*DBN Deep - Belief Network (Deep Learning)
*DNN-HMM - Deep Neural Network-Hidden Markov Models (Hybrid Deep Learning)
*DNN-CRFs - Deep Neural Network-Conditional Random Fields (Hybrid Deep Learning)
*CDNN-RNN - Convolutional Deep Neural Network- Recurrent Neural Network (Hybrid Deep Learning)

NA- Not Applicable