**A COMPREHENSIVE REVIEW OF DEEP LEARNING ARCHITECTURES FOR TASK SPECIFIC ANALYSIS**

**Minor Project-II**

**(ENSI252)**

*Submitted in partial fulfilment of the requirement of the degree of*

**BACHELOR OF TECHNOLOGY**

*to*

**K.R Mangalam University**

*by*

**Sahil Bhardwaj (2301010425)**

**Rishav Kumar (2301010424)**

**Ayush Verma (2301010447)**

Under the supervision of

**Dr. Solanki Gupta**

**<Internal>**

**Associate Professor**

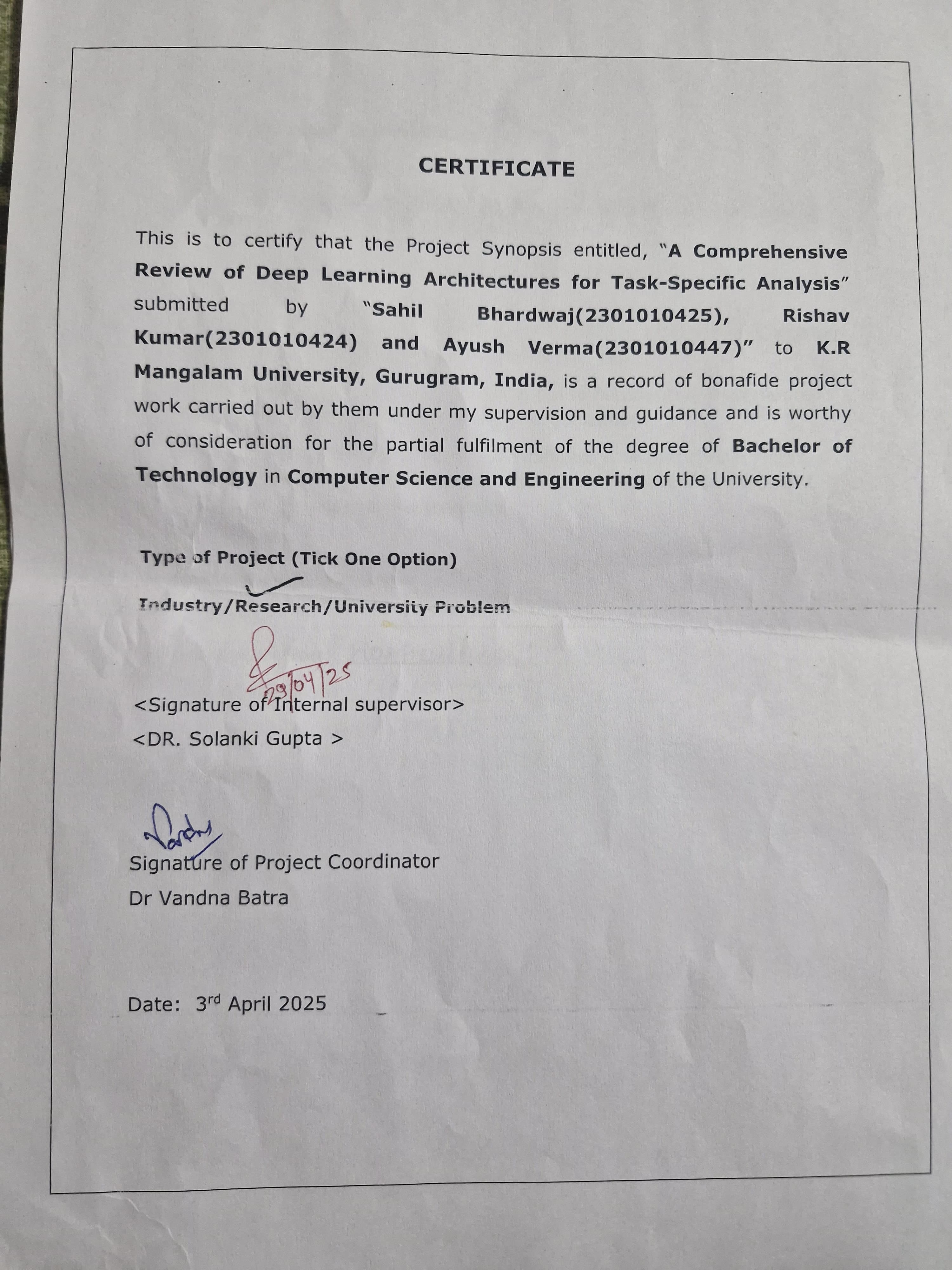


Department of Computer Science and Engineering

School of Engineering and Technology

K.R Mangalam University, Gurugram- 122001, India

April 2025



**INDEX**

|  |  |  |
| --- | --- | --- |
|  | Abstract | Page No. |
|  | Introduction (description of broad topic) |  |
|  | Motivation |  |
|  | Literature Review/Comparative work evaluation |  |
|  | Gap Analysis |  |
|  | Problem Statement |  |
|  | Objectives |  |
|  | Tools/platform Used |  |
|  | Methodology |  |
|  | Experimental Setup |  |
|  | Evaluation Metrics |  |
|  | Results And Discussion |  |
|  | Conclusion & Future Work |  |
|  | References |  |
|  | Annexure I  Research Paper (Published/Submitted) |  |

**ABSTRACT**

Deep learning has truly changed the game across numerous fields, reshaping how we tackle complex challenges by providing highly precise and efficient solutions tailored to particular needs. Just picture a system that can create text, summarize information, translate languages, classify data, answer questions, and even reason— deep learning makes all of this a reality. In this review, we took a closer look at different deep learning architectures and see how they drive these various applications. We analysed the past studies and reveal the datasets that power these models, as well as the design principles that influence their performance. Throughout this we emphasized the strengths that set these architectures apart, along with the limitations that pose challenges to their effectiveness. This review acts as a guide for researchers, practitioners, and industry professionals, helping them choose and adapt the right deep learning models for specific tasks. By presenting insights into both established and emerging deep learning trends, this review serves as a guide for advancing the development of more robust, interpretable, and generalizable deep learning systems.

***KEYWORDS: Deep Learning, Deep Learning Architectures, Task Specific Review, Systematic Review***

**Chapter 1**

**Introduction**

1. **Background of the project**

Deep learning is a subfield of machine learning and artificial intelligence that attempts to model the way humans learn from information to extract patterns for decision-making. Neural networks that have multiple layers are used to deal with large data. Through these layers, deep learning models can learn complicated patterns and representations, making them efficient for many applications, including image recognition, natural language processing (NLP), and speech recognition. The very idea that deep learning embodies is that it allows machines to learn directly from data in their raw form, such as an image with its associated text or audio, without human intervention in feature extraction. This process favours neural networks made up of nodes or interconnected neurons, which adjust their weight and biases during training to minimize errors and increase accuracy. What holds the transition from classical machine learning to deep learning is that the classical machine is doing manual feature extraction, while deep networks are learning directly from raw data. Classical machine learning works great with small datasets but often struggles to infer on complicated patterns, while deep learning works exceptionally well with large datasets, achieving reasonable accuracy for image recognition and NLP. On the other hand, deep learning does require the use of Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs) for processing power, whereas traditional machine learning could run just fine on standard Central Processing Units (CPUs). Traditionally machine learning methods are applied to structured data, while deep learning suits unstructured data applications. This evolution has come into place under the influence of computing power becoming inexpensive, the existence of big data, and user-friendly frameworks such as TensorFlow and PyTorch. Task-specific analysis in deep learning is important since tasks have individual architectural needs for achieving the best performance. Text generation, for instance, necessitates contextual coherence and flow, so Transformers [99] such as Generative Pre-trained Transformer (GPT) are best suited because they process sequential data. Summarization is about extracting the gist of a text, where sequence-to-sequence models with attention such as Text-to-Text Transfer Transformer (T5) or Bidirectional and Auto-Regressive Transformers (BART) are used to emphasize significant input segments. Translation needs precise language mapping, which is a strength of encoder-decoder architectures. Classification is aided by less complex architectures such as Convolutional Neural Networks (CNNs) or fine-tuned Bidirectional Encoder Representations from Transformers (BERT) models for effective feature extraction. Question answering (QA) requires contextual awareness to provide accurate answers, utilizing models such as BERT with attention. Reasoning requires logical conclusions and multi-step processes, necessitating sophisticated models such as GPT-4 with memory layers. Adapting architectures to task requirements provides improved performance and more accurate outcomes. Deep learning applications in any specific task encounter challenges such as limited availability of data and lack of quality annotation, which leads to problems in model training and generalization. While complex architectures prevent overfitting at times, highly skilled regularization may be demanded. These models are often black boxes, making them hard to interpret; interpretability is critical for sensitive tasks like healthcare. Generalization is difficult, requiring intense fine-tuning and transfer learning. Computational demands are high, thus increasing the cost and energy. Other ethical concerns include the biases embedded in them, which may lead to failure in achieving fair outcomes. Real-time tasks face problems caused by latency, making their deployment in the interactive environment harder. Besides, all these challenges require model design, data preparation, and constant monitoring to be addressed. A review paper that systematically takes into account these questions, datasets employed, the rationale behind their design, and the pros and cons of various models would provide valuable insights into problems related to task-specific deep learning applications. Therefore, the major objectives of this study are to explore:

* The deep learning architectures that are most commonly used across different tasks.
* The datasets utilized in these studies and the principles behind their design.
* The strengths and weaknesses of various models.

To start the review, we identified key real-world applications of deep learning, including text generation, text classification, reading comprehension, summarization, reasoning, translation, and question answering, as fundamental tasks for analysis. By looking at the functionality and performance of model architectures with respect to these tasks, this review would help us understand model complexity and overfitting issues and reach suggestions for reasonable regularization strategies. An evaluation of strong and weak points of different models would help in proposing interpretable and generalizable architectures by reducing the behavior of a deep learning model as a black box. It would also offer guidance for data quality and availability by addressing the most efficient datasets and consequent drawbacks, thereby leading to better strategies for dataset selection and augmentation. Further, this review study would serve as the perfect guide for understanding the efficient model for different task-specific deep learning applications.

1. **MOTIVATION**

The explosive growth of deep learning has led to remarkable achievements across various domains, yet selecting the right architecture for a specific task remains a significant challenge. Different tasks such as text generation, classification, summarization, translation, question answering, and reasoning each demand unique architectural designs and training strategies to achieve optimal results. Despite the abundance of deep learning models, there is often a lack of consolidated guidance to help researchers, practitioners, and industry professionals understand which models perform best for which tasks, why certain architectures succeed or fail, and what trade-offs are involved.

Moreover, while many studies focus on introducing new models, few provide a holistic comparison that systematically analyzes existing architectures in the context of specific applications. Real-world problems also pose additional challenges like dataset limitations, high computational requirements, model interpretability, and ethical concerns such as bias and fairness, which are not adequately addressed in fragmented literature.

This paper was motivated by the need to bridge this gap by conducting a comprehensive review and comparative analysis of deep learning architectures for task-specific applications. By highlighting strengths, limitations, dataset choices, and design principles, this study aims to serve as a practical guide for informed model selection, fostering more efficient, interpretable, and ethical application of deep learning in real-world scenarios.

**Chapter 2**

**LITERATURE REVIEW**

1. **Review of existing literature**

Deep learning has evolved into a foundational technology enabling numerous real-world applications across industries like healthcare, finance, education, and autonomous systems. In recent years, extensive research has been conducted to improve deep learning models for task-specific applications, leading to the development of various architectures tailored to different domains.

Several works have explored deep learning models for reading comprehension, where initial methods relied on shallow machine learning techniques such as Support Vector Machines and Decision Trees. However, the advent of models like BERT, ALBERT, and T5 introduced transformer-based architectures that significantly enhanced the contextual understanding of text, achieving state-of-the-art results in datasets like SQuAD and ReClor.

In the area of text generation and summarization, early methods such as rule-based and statistical techniques (e.g., TF-IDF and keyword extraction) showed limitations in handling semantic meaning and coherence. Modern approaches, particularly sequence-to-sequence models with attention mechanisms and Transformer-based models like GPT-2, BART, and PEGASUS, have enabled the generation of fluent, contextually relevant summaries and creative text generation.

For translation tasks, traditional rule-based and statistical machine translation systems have been largely replaced by Neural Machine Translation (NMT) models. Architectures like Deep Transition RNNs, Vanilla Seq2Seq, and Transformer models have shown significant improvements, with Transformer variants dominating benchmarks like WMT'14 English-German tasks due to their capability to model long-range dependencies without recurrence.

In classification and question answering (QA) tasks, fine-tuned pre-trained models like BERT and RoBERTa have demonstrated exceptional performance, offering superior feature extraction and contextual reasoning. Hybrid approaches combining retrieval-based and generative models have further enhanced QA systems, particularly in complex, open-domain settings. Research into reasoning tasks using deep learning has also progressed, although challenges remain. Models like GPT-3, Codex, and Minerva have demonstrated the ability to perform multi-step reasoning, but gaps in consistency, robustness, and true logical reasoning still persist. Traditional logical systems and symbolic reasoning have been integrated with neural approaches to address these limitations, resulting in hybrid neural-symbolic models that attempt to combine the benefits of both paradigms.

Despite these advancements, common challenges across domains include:

* High computational costs.
* Model interpretability and transparency issues.
* Difficulty in generalizing across tasks and domains.
* Ethical concerns related to bias and fairness embedded in models.

Several surveys have discussed these challenges individually across domains; however, comprehensive studies comparing deep learning architectures across multiple task-specific applications, highlighting their strengths, weaknesses, and dataset dependencies, remain limited.

Thus, this paper positions itself to fill that gap by systematically reviewing deep learning architectures across major tasks such as generation, translation, classification, summarization, question answering, and reasoning, providing critical insights for researchers and practitioners seeking task-optimized deep learning solutions.

**Table 1**: Summary of Popular Studies on Reading Comprehension Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Datasets | Key takeaways | Limitations | References |
| BERT, RoBERTa  DistilBERT, ALBERT | ReClor | 1. ALBERT is the best model in this paper, so far.  2.Polytuplet loss improves accuracy by 5.6%-11.7% over baseline models like ALBERT, BERT, and DistilBERT. | 1. The comparison is limited to baseline models, without evaluating techniques.  2. Scalability to large datasets or real-world tasks remains unaddressed. | **[5]** |
| T5 base model  BART base model  GPT-2 model | Fairytale QA Corpus  Textbook Question Answering (TQA) dataset | 1. The paper compares different neural architectures for automatic question generation based on reading comprehension passages  2.Highlights the strengths and weaknesses of various Question Generation models. | 1. The paper focuses on a narrow set of models, lacking comparison with a broader range of question generation techniques.  2. The evaluation metrics used might not fully capture the complexity of question quality. | **[6]** |
| Stanford AR  GA Reader | Who-Did-What (WDW)  Children’s Book Test (CBT) | 1. The paper compares word embedding techniques like GloVe, Word2Vec and fastText for reading comprehension tasks.  2.Embedding effectiveness varies based on task and dataset. | 1.Focus is less on word embedding models.  2. The paper doesn’t provide detailed insights into why some embeddings perform better than others. | **[7]** |
| CNN & Daily Mail, CBT, Who-did-What, CLOTH, CliCR MCTest, RACE, LAMBADA  SQuAD, NewsQA, TriviaQA, DuoRC  BERT, GPT  RNN  CNN | CNN &Daily Mail  CBT  -LAMBADA  -CLOTH  - RACE  - SQuAD | 1.The paper proposes a neural network-based model that reads and understands a passage to answer questions without requiring task-specific feature engineering. 2. It utilizes attention mechanisms to focus on relevant parts of the text, improving the model's comprehension ability. | 1.The model's achievement heavily relies on the quality and size of the training data, which may limit its effectiveness in low-resource settings.  2.The model requires significant computational resources, especially for large-scale datasets and complex attention mechanisms. | **[8]** |
| BERT  RoBERTa  Cross-Document Reasoning Models  Textual Entailment Models | TriviaQA Web  DuReader | 1. The paper introduces a model that performs reading comprehension across multiple documents, capturing information from diverse sources to answer questions.  2.Attention mechanisms have been used to attention the model's reasoning process at the most relevant parts of the files. | 1.The method includes complicated architectures that can be computationally expensive, requiring sizable resources for schooling and inference, particularly with huge file sets.  2. The model doesn't provide fine-grained control over which documents or pieces of information are prioritized in the reasoning process. | **[9]** |
| Co-match  BERT | MCTest  CNN/Daily Mail  RACE | 1.BERT performed higher accuracy compared to the Co-match model on the Vietnamese corpus.  2.It targets multiple-choice reading comprehension questions, where the model selects the most appropriate answer based on the given passage. | 1.The study is tailored to Vietnamese, limiting its applicability to other languages with different linguistic structures.  2.The paper primarily compares a few deep learning models (RNNs, LSTMs, and BERT), without considering a broader range of models or alternative architectures. | **[10]** |
| T5  BERT | DROP | 1.The proposed method demonstrates significant improvements over baseline models, achieving higher F1 scores on the hard subset of the DROP dataset.  2. A single model is used for both question decomposition and reading comprehension, simplifying the architecture. | 1. Performance is reliant on the availability of annotated sub-questions, and weak supervision can only partially alleviate the data limitation.  2. The success of the model relies on the accuracy of the question decomposition process, which remains a challenging task. | **[11]** |
| Bi-GRU  Encoder | CNN/Daily Mail | 1.Improves query-document interaction for improved answer selection.  2.Surpasses state-of-the-art models on CNN/Daily Mail and CBTest datasets. | 1. The system can misunderstand ambiguous requests.  2.Attention mechanisms incur computation cost. | **[12]** |
| LSTMs | SQuAD | 1. Splits MRC into Cloze-fashion, multi-preference, span-prediction, and free-form question answering.  2. Pre-educated models (BERT, GPT,) outperform baseline strategies in contextual comprehension. Knowledge. | 1.Difficulty in dealing with lengthy text passages, resulting in loss of contextual pertinence.  2. Needs huge-scale labelled datasets to prevent overfitting. | **[13]** |
| GPT-FT | COSMOS QA | 1. The paper introduces a model that integrates contextual commonsense knowledge to improve machine reading comprehension, enabling better understanding beyond explicit information in the text.  2. The model tailoring knowledge application to the specific reading passage and question. | 1. The model might struggle to generalize to very diverse or uncommon knowledge that is not well-represented in the commonsense knowledge base.  2. Integrating contextual commonsense reasoning adds computational complexity, which can slow down training and inference times. | **[14]** |

. **Table 2**: Summary of Popular Studies on Translation Task

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | Datasets | Key Takeaways | Limitations | References | |
| Deep Transition RNNs  Stacked RNNs Bi-Deep RNN Architecture | WMT'15 English-German (En-De)  Byte Pair Encoding (BPE)  WMT'14 English-French (En-Fr) | 1.The research proves that extra profound architectures, specifically deep transition RNNs and stacked RNNs, decorate neural gadget translation (NMT) accuracy.  2.Increasing version intensity (up to eight layers) assists in taking snap shots greater state-of-the-art linguistic styles, enhancing translation accuracy. | 1.Training deep NMT models is time-consuming and computationally steeply-priced, making them much less scalable for large-scale programs.  2. Although the consequences are encouraging, the generalizability of the method to different language pairs or domain names is but to be hooked up. | **[18]** | |
| Vanilla Seq2Seq Model  RNNs  GRUs  LSTMs | WMT’15 English-German Task | 1.Using deeper fashions, mainly with interest mechanisms, leads to seriously better translation excellent.  2.The models showed study overall performance across both English-German and English-French, confirming their versatility. | 1.The Transformer and bidirectional models required longer education times.  2.Taken a look at frequently specializes duties in English-German and English-French translation duties, which can be every pretty excessive-resource language pairs. | **[19]** | |
| Transformer Models  DLCL | WMT'16 English-German (En-De) Task  NIST'12 Chinese-English (Zh-En-Small) Task  WMT'18 Chinese-English (Zh-En-Large) Task | 1.Deepening the Transformer model results in improved translation quality, particularly when utilizing dynamic layer combinations (DLCL).  2.Employing a dynamic combination of layers in the encoder and decoder enhances translation quality over conventional fixed-layer models. | 1.Deep models such as Transformer-Deep are highly computationally demanding and may prove difficult to train on regular hardware because of processing and memory constraints.  2.Because of the model's depth, it is hard to process a complete batch on one GPU. | **[20]** | |
| DTMT  Vanilla Encoder-Decoder  Transformer Model | WMT'14 English-German (En-De)  IWSLT'15  Multi30k | 1.Bloating the Transformer model results improves translation, particularly when utilizing dynamic layer combinations DLCL.  2. Using a dynamic combination of layers within the encoder and decoder pairs well with translation spectacularity compared to traditional constant-layer models. | 1.Deep models such as Transformer-Deep can be challenging to train on regular hardware because of processing and memory limitations.  2.Due to the version's severity, it's miles hard to prepare a complete set on a solitary GPU. | **[21]** | |
| Vanilla Encoder-Decoder Models  BERT  GPT  CNNs  LSTM  GRU | WMT  IWSLT  Flickr30k and COCO | 1. Pretrained models can be transferred to NMT tasks and thus are very effective for low-parallel-data languages.  2.Pretraining language models like BERT and GPT has been shown to enhance performance on the task of translation. | 1.NMT models need large data sets to effectively train.  2.NMT models can inherit and pass on biases in the training data, resulting in biased or unfair translations. This is a major issue in applications where fairness and neutrality are important. | **[22]** | |
| Seq2Seq  Transformer Models | American Sign Language (ASL) | 1.These pairs scored better than different styles, with better BLEU rankings on the GSL dataset.  2.Higher models proved robust ability on much less controlled ASL and CSL datasets, showing versatility. | 1.Performance can fluctuate with less managed facts units because of variability in signing patterns and recording environments.  2.Advanced models, such as transformers, necessitate substantial computational resources throughout both the training and inference phases. | **[23]** | |
| Neural Machine Translation | WMT (Workshop on Machine Translation) | 1.Translation Adequacy: In blind tests, CUBBITT performed better than professional human translators in maintaining the original meaning of the text.  2.Fluency Comparison: Human translations were graded as more fluent. | 1.Fluency Gap: There is still a narrow fluency gap between CUBBITT's outputs and those of human professionals.  2.Domain Specificity: The performance of the system has been mostly tested on news articles, and its performance on other domains or language pairs might need to be evaluated. | | **[24]** |
| Global Memory Module | IWSLT | **1.**The model presented here greatly enhances translation quality by efficiently capturing and making use of both local and global context information.  **2.** Integrating grammatical dependencies with the attention mechanism enhances context representation, resulting in more precise translations**.** | 1. The "end-to-end" design of deep learning models may result in poor interpretability of learning outcomes, making it hard to know the decision-making process.  2. Although the model is good on the IWSLT dataset, its generalization to other datasets or real-world use needs to be verified. | | **[25]** |
| RBMT and SMT | United Nations Parallel Corpus | 1.The move from rule-based and statistical models to neural models has tremendously improved the quality of translations.  2.Combination of various MT paradigms can effectively cope with particular issues, like low-resource languages | 1.NMT systems need huge quantities of good quality parallel data, and such parallel data may not exist for all language pairs.  2.Even with progress, getting high-quality translations for low-resource languages is still a major issue. | | **[26]** |
| Transformer-Based NMT | Translation Corpus (TC) | 1.The multi-challenge mastering method enhances MAP by using 16% in comparison to the baseline transformer.  2.Evades overfitting to TC terminology, producing  translations relevant to each corpora. | 1.The model works well but keeps quite low MAP rankings as a result of having few education epochs and dataset.  2. Speaks to gaining access to a retrieval corpus (RC) index, which hinders schooling index. | | **[27]** |

**Table 3**: Summary of popular Studies on Summarization Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Data Sets | Key takeaways | Limitations | References |
| STFIDF  TBS | BillSum,  IN-ABS and  IN-EXT | 1.Legal structures vary by areas, so there is a need for models evolved on jurisdiction-particular statistics to stay accurate and relevant.  2.With prison documents written in diverse languages, institutions including the European Union, there may be growing call for for fashions able to doing multilingual and pass-lingual summarization obligations. | 1.There is a great scarcity of huge-scale, outstanding datasets across most jurisdictions and languages, making it hard to build sturdy legal summarization fashions.  2.Traditional assessment metrics may fail to effectively capture actual correctness and legal soundness of summaries, that are vital in prisoneventualities**.** | **[31]** |
| RNN Extractor and Seq2Seq Extractor  Cheng & Lapata Model | Reddit and  AMI | 1.Position Bias: Sentence function is the main responsibility that summarization models must bear.  2.Word averaging is just as good as CNNs/RNNs. | 1.Performance isn't always consistent throughout domain names.  2.Models are prone to overfitting dataset-specific characteristics. | **[32]** |
| GoogleNet and AlexNet  LSTM | YouTube | 1.Deep learning outperforms conventional approaches (CNNs, RNNs, Transformers enhance summarization).  2.Supervised models are precise but require large labelled datasets (SumMe, TVSum). | 1.Prevention of duplicate or  Unwanted files.  2.Preservation of meaningful and  contextually appropriate segments. | **[33]** |
| Seq2Seq | DUC (Document Understanding Conferences) | 1.Getting to know that strategies have dramatically progressed the overall performance of MDS systems.  2.The authors advocate a brand-new taxonomy classifying neural community design methods for MDS. | 1.Super datasets required for powerful training of deep mastering models.  2.Deep studying algorithms for MDS tend to call for loads of computational assets, consequently less suitable for researchers with confined facilities. | **[34]** |
| RNNs  and  BERT SUM, T5, PEGASUS | Gigaword | 1.It discusses the evolution of models from RNNs and LSTMs to more advanced transformer models, showing improvements in generating coherent and concise summaries.  2.ROUGE-1, ROUGE-2, and ROUGE-L are the maximum broadly used metrics to evaluate summarization pleasant | 1.Traditional evaluation metrics like ROUGE may not fully capture the quality of abstractive summaries, especially when it comes to factual accuracy and coherence.  2.The models often face challenges when dealing with out-of-vocabulary (OOV) words, which can negatively impact summary quality, especially in specialized domains | **[35]** |
| Attention Mechanisms | Pre-Training and Fine-Tuning | 1.The method below attention utilizes deep fashions that are trained extensively on big datasets through pre-training and excellent-tuned with domain-particular net pages.  2.The technique indicates area adaptability, effectively moving to extraordinary net domain names by way of exceptional-tuning pre-trained fashions with little domain-unique facts. | 1.The fulfilment of the approach is largely dependent on the presence of first rate, large-scale datasets for pre-training and satisfactory-tuning.  2.Although the technique contains extraordinary fields, a few specialized domains with precise terminologies or frameworks may want in addition adjustment. | **[36]** |
| GPT-2  BERT | Udacity Lecture Transcripts | 1.BERT plays higher than traditional tactics in summarizing lectures.  2.K-Means clustering lets in for dynamic adjustment of summary duration in line with consumer desire. | 1.Difficulty with prolonged lectures (a hundred sentences may lose context).  2.Computationally highly priced (BERT could be very useful resource-in depth). | **[37]** |
| Coverage Models | Gigaword | 1.Integration of attention mechanism and pointer-generator network has enormously enhanced the generated summary's quality.  2.Having access to large and high-quality datasets is imperative for training good summarization models. | 1. Abstractive models have the possibility of creating information that does not exist in the source material, creating possible inaccuracy.  2. Advanced deep learning models take a lot of computational resources, and this may not be readily available to all researchers**.** | **[38]** |
| Graph-Based Methods  Template-Based Methods | TAC (Text Analysis Conference) | 1.Dependent on the selection of pre-existing sentences; While using is less complex to use, it can be repetitive and incompatible.  2.Creates new sentences that forms the content of the text; Greater is flexible but more difficult because it asks for herbal language era's abilities. | 1.It may not be readable to see that they can be based on literal sentences, which can also bring about excesses and lack of glide.  2.Sophisticated natural language production strategies and a large amount of education fabrics require; They are also interrupted with the help of problems in preserving the data up-to-date | **[39]** |
| BERT  BiGRU | IMDb Reviews | 1.Model design selection must conform to the inherent nature of the text classification task and consider the size of the sequence and the significance of the reference. | 1.It takes a large amount of computational resources, especially to train and inferior transformer-based architecture.  2.Model can overfit training data, especially when working with small datasets. | **[40]** |

**Table 4**: Summary of popular Studies on Question-Answering Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Datasets | Key takeaways | Limitations | References |
| BERT  Hybrid Models | MEDIQA | 1.Deep mastering has immensely improved the performance of scientific QA systems to recognize and bring natural language more efficiently.  2.Blending similar approaches, such as retrieval-based solely and understanding-based solely models, is likely to yield improved outcomes than a single approach. | 1.Medical vocabulary and the complexity of medical language are challenging for herbal language processing fashions.  2.It is difficult to quantify overall performance of medical QA systems because medical recommendation is subjective and there may be variations in correct solutions**.** | **[47]** |
| Dataset-Specific Optimized Models  Inter-Sentences Architecture | TREC QA | 1.Architectures that model question-answer interactions at earlier stages (word or subsquence level) work better.  2.The paper brings to the attention that different architectures provide different performance, and there is a focus on the correct choice based on the application to be addressed. | 1.The performance is only measured for the TREC QA dataset and therefore may restrict the generality of the results to other datasets or domains.  2.Four provided architectures comprise the scope of the research, although potential models different from them are not taken into account**.** | **[48]** |
| FCNs and  LSTM | PASCAL VOC | 1.Deep learning algorithms have greatly enhanced the accuracy of semantic segmentation operations compared to conventional techniques.  2.The encoder-decoder architecture, i.e., the network architecture, is crucial in maintaining the equilibrium between localisation accuracy and context capture. | 1.A model can be trained on a particular dataset but can be poor in another scenario or domain.  2.Model will require domain adaptation methods or other training sets. | **[49]** |
| Autoencoders  DBNs | Electronic Health Records | 1.Deep learning algorithms have much enhanced the accuracy of disease diagnosis and prognosis.  2.Deep learning allows heterogeneous data sources to be integrated, offering an integrated view of patient health. | 1.Utilization of sensitive patient information poses concerns related to privacy and security**.**  2.The accuracy of these models highly relies on the quality as well as the availability of data, which in healthcare applications can prove to be a limiting factor**.** | **[50]** |
| Information Retrieval and Deep Neural Network | WikiQA | 1.The discipline has moved from the classical IR-based approach to integrating deep learning methods, resulting in huge leaps in comprehending and creating correct answers.  2.Merging IR and DNN techniques has the potential to capitalize on both approaches. | 1.Deep learning model training and deployment require immense computational resources, which might be out of reach for some organizations. | **[51]** |
| GRU  Dynamic Memory Networks | MCTest Dataset (Microsoft) | 1.Attention-based models assist in extracting information pertinent to answering questions.  2.Sequence-to-sequence models work well to produce multi-word responses. | 1.Training certain models, such as Dynamic Memory Networks, is computationally costly.  2.Performance is constrained by fixed memory sizes on long-context tasks. | **[52]** |
| GANs | Genomic Databases | 1.Deep learning algorithms can enhance the accuracy of disease detection and diagnosis.  2.Models allow for customized treatment protocols based on specific patient information. | 1.The management of sensitive patient data requires strict privacy practices.  2.Deep learning models tend to be "black boxes" and difficult to interpret their decision-making processes. | **[53]** |
| MRC | SQuAD (Stanford Question Answering Dataset) | 1.The incorporation of deep learning methods, particularly neural networks, has greatly enhanced the performance of open-domain QA systems.  2.A range of models, such as MRC, knowledge-based, and hybrid models, serve various aspects of QA tasks, and the choice of suitable models depends on the application. | 1.Certain models struggle to scale to large datasets or process the enormous amount of information present in open-domain environments.  2.Models can still be challenged by grasping subtle contexts or unclear questions and provide the wrong answers**.** | **[54]** |
| TPRN | SQuAD | 1.TPRN encodes grammar-like structures without explicit annotation.  2.Symbol-role binding enhances readability by linking words with grammatical functions. | 1.Lower accuracy than BiDAF (~2% loss of F1 score).  2.Takes large computational power for training and tuning. | **[55]** |
| SGD  Elastic Averaging SGD | TREC QA | 1.Distributed deep learning speeds up training procedures.  2.Optimization algorithm performance is inconsistent; whereas certain ones such as EASGD perform well under distributed environments | 1.Although improved, the speedup from increased workers is sublinear, which means returns diminish as more workers are added.  2.Distributed training brings communication overhead, which can negate the advantages of parallelism, particularly in high-latency environments. | **[56]** |

**Table 5**: Summary of popular Studies on Generation Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Datasets | Key takeaways | Limitations | References |
| GANs | Speech Data | 1.The usage of GANs is to generate new and unique architectural shapes, going past traditional design methods.  2.The article suggests a graph-primarily based gadget learning approach for 3D architectural layout spaces. | 1.Mapping 3D architectural designs into graph representations can be challenging.  2.Deep neural network training, particularly GANs, on graph-based facts calls for excessive levels of computation. | **[60]** |
| VAEs | Synthetic Data Output | 1.The proposed structure, SenseGen, generates synthetic sensor statistics, permitting information argumentation and device mastering version education.  2.Employing deep studying models, SenseGen analyses state-of-the-art patterns in sensor readings. | 1.The validity of synthesized records generated closely depends on the domain the original information comes from.  2.It continues to be challenging to assess the pleasant and usability of artificial statistics. | **[61]** |
| RNNs Autoencoders | Audio-Based Datasets | 1.Deep learning can facilitate diverse purposes like melody composition, polyphony, accompaniment, and counterpoint.  2.Generation strategies such as single-step feedforward processes, iterative feedforward strategies, sampling strategies, and input manipulation are employed to control the music generation process**.** | 1.The diversity-coherence trade-off in generated music is still an issue.  2. The majority of current models cannot integrate real-time user feedback to any extent. | **[62]** |
| Point Cloud-Based Models and  Voxel-Based Models | ShapeNet and ModelNet | 1.Deep learning has greatly improved the ability to create complex and sundry 3-D shapes out of reach of conventional modelling.  2.The survey categorizes cutting-edge models into several classes, giving a scientific evaluation of the methodologies in the field. | 1.Computational models based totally on deep learning for the technology of 3-D form require substantial computational sources, which may not be less costly for everybody who's either a practitioner or a researcher.  2.The models won't generalize to new instructions of shapes, proscribing their application | **[63]** |
| LSTM and  Hybrid Models | Stock Market Data and Electricity Consumption Dataset | 1.Deep models including RNNs and LSTM networks are proven to outperform conventional statistical models in the modelling of sophisticated temporal dependencies.  2.Blending deep learning models with classical forecasting techniques or other machine learningtechniques can result in better forecasting accuracy and stability. | 1.Deep learning models need vast amounts of high-quality training data, which in time series applications may not always be available**.** 2. Deep learning models, if not properly regularized and validated, can overfit with small datasets**.** | **[64]** |
| RNNs and CNNs  SaShiMi | Music Generation and  Unconditional Speech Generation | 1.It integrates S4 layers with a multiscale structure to enable efficient modelling of long-range dependencies in audio data.  2.Resolves S4 autoregressive generation stability by modifying parameterization, keeping it stable while generating audio. | *Autoregressive Instability***:** The standard S4 models are unstable during autoregressive generation and need to be parameter-tuned**.** | **[65]** |
| VAE  MedGAN | MIMIC-III & Sutter EHR | 1.VAEs and GANs are state-of-the-art methods in artificial data generation.  2.Synthetic privacy-preserving data ensures secure data sharing. | 1.Models learn biases from actual datasets.  2.Excessive resource utilization for training generative models. | **[66]** |
| LSTM | ImageNet | 1.Chainer introduces "Define-by-Run" execution, making deep learning models more flexible and easier to use.  2.Optimized GPU computation using CuPy for speeding up deep learning training. | 1.Models will be prone to inherit real-world dataset biases.  2.Vast resource demand for training generative models. | **[67]** |
| BERT  RoBERTa | PY150, GitHub | 1.CodeXGLUE has 14 datasets for 10 programming tasks such as code search, code translation, and bug detection.  **2.**Integration of pretrained models: uses CodeBERT, CodeGPT, and Encoder-Decoder as baselines. | *Not having Real-World Edge Cases: Certain* datasets are generated synthetically and lack actual real-world variations in coding. | **[68]** |
| RNN | WikiText-2 | *Inclusion of Recurrent Neural Networks (RNNs):*The bigger architecture includes RNN layers within the Transformer model, in the hope of better capturing sequential relationships. | *Risk of Overfitting:* The bigger model, with its greater number of parameters, can be more overfitting, particularly when dealing with smaller datasets**.** | **[69]** |

**Table 6**: Summary of Popular Studies on Reasoning Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Dataset | Key takeaways | limitation | References |
| GPT-3, PaLM, Codex, RoBERTa, T5, Transformer, Seq2Seq and Minerva, GPT-3, MWP-BERT, Bhaskara, NaturalProver, UniGeo, FinQANet | **MathQA**, **CoqGym**,  **GEOS** **TheoremQA**  **ScienceQA** | **1.GPT-3, PaLM (Minerva), Codex have shown advanced reasoning talents.**  **2.GPT-3, Minerva, MWP-BERT, Codex carry out properly but aren't optimized for math reasoning.** | 1.Pre-educated Language Models Are Not Optimized for Math Reasoning.  2.Lack of Consistency and Robustness in Mathematical Reasoning | **[75]** |
| BERT RoBERTa  T5  GPT-3  T5-11B RoBERTa  GPT-3 | DeepMind Mathematics Dataset  SVAMP  **HOList**, **ParaRules** | 1.Transformers Achieve High Performance however, Lack True Reasoning.  2.Models like BERT, GPT-three, RoBERTa, and T5 carry out properly on many NLP tasks however battle with deep reasoning. | 1.Math Word Problems (MWPs) are challenging, and performance drops when questions are barely changed (e.g. SVAMP dataset).  2.Fail on lengthy sequences requiring reminiscence. | **[76]** |
| BERT, T5, RoBERTa  Graph Neural Networks (GNNs), Relational Networks Memory-Augmented Neural Networks (MANN) | **CLEVR**  ATOMIC **DeepMind Mathematics Dataset** | 1.Neural networks like Transformers (BERT, T5, RoBERTa) are good at learning statistical patterns in data.  2.They lack deliberate logical reasoning and fail to systematically generalize beyond their training data. | 1. Deep learning is still at the surface level (lacks true reasoning).  2.Neural networks rely on statistical correlations rather than true logical deductions. | **[77]** |
| **RTNs & RNTNs** | **DBpedia** | 1.Classical logic-based reasoning is correct but slow and does not handle incomplete information well.  2.Relational Tensor Networks – RTNs can substitute rule-based reasoning for faster and more scalable ontology inference. | **1.RTNs do not guarantee strict logical correctness like traditional reasoning systems.**  **2.RTNs struggle with nested logical rules, negation, and deep inference chains.** | **[78]** |
| CNN, LSTM, Transformer and EDNNs, FDNNs, RDNNs | **MNIST Datasets** MedicalAI Datasets **Cybersecurity Datasets (e.g., CICIDS2017, NSL-KDD)** | 1.Traditional notion/proof theories were used for reasoning below uncertainty.  2.Belief Theories Can Enhance Deep Learning Models and Three varieties of uncertainty-aware deep studying fashions mentioned. | 1.Handling noisy or opposed statistics stays a mission, specifically in hostile assaults on AI models.  2.Uncertainty estimation methods can extend biases if not cautiously designed. | **[79]** |
| CNN, MLP, LSTM and EDNNs, FDNNs, RDNNs | **Sandia Matrices, RAVEN-FAIR** PGM | 1.Understanding uncertainty is fundamental to effective selection-making in AI and deep mastering.  2.Fuzzy Deep Neural Networks (FDNNs) – Uses Fuzzy Logic for indistinct data and Rough Deep Neural Networks (RDNNs) – Uses Rough Set Theory to version imprecise or incomplete records | 1.Deep studying models overfit unique RPM systems rather than gaining knowledge of real abstract reasoning.  2.PGM dataset exposes generalization disasters, as many models fail on feature distributions. | **[80**] |
| Transformer | **Multiple VQA datasets**, | 1.Transformer-based approach that enhances visual reasoning through self-attention and co-attention mechanisms model iteratively refines its understanding of images and text.  2.Using custom tokens improves how the model integrates visual and textual features for better comprehension. | 1.The model struggles slightly with counting and numerical comparison tasks, achieving lower accuracy in "Compare numbers" type question on the CLEVR dataset. 2. While Transformers excel in capturing relationships, they are computationally expensive and require high-end hardware for training and inference. | **[81]** |
| Hybrid Neural-Logical Model | Sudoku Datasets Protein MPNN dataset | 1.Deep studying with logical reasoning permits solving NP-tough issues more effectively.  2.E-NPLL loss overcomes barriers of conventional pseudo-loglikelihood features, enabling better logical constraints. | 1.While the approach is scalable, fixing very big NP-hard problems with many variables nevertheless poses computational stressful conditions, especially in the course of inference.  2.The E-NPLL loss, however deciding on the proper good enough charge (wide sort of omitted variables) is crucial, affecting convergence tempo. | **[82]** |
| **RNNs,** LSTM | Kinsources, IMDb,CLUTRR (Commonsense Reasoning Benchmark | 1.The paper offers a hybrid version that combines Neural Networks with First-Order Predicate Logic for analogical reasoning.  2. Analogical Reasoning Outperforms Traditional Deductive and Inductive Methods. | 1.It struggles while dealing with large-scale understanding bases with hundreds of thousands of facts.  2.The model is quite specialized for logical inference and dependent symbolic responsibilities, making it less effective for unstructured records. | **[83]** |
| **CNN, RNN** | Clevr | 1.CBN-based models beat humans on CLEVR (97.6% accuracy).  2.Acquire multi-step reasoning without being taught explicit compositional structure. | 1.Struggling with longer, more involved problems.  2.There is no direct hierarchical modelling, as compared to domain-specific architectures. | **[84]** |

**Table 7**: Summary of Popular Studies on Classification Task

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Datasets | Key takeaways | Limitations | References |
| LSTM, GRU  Hybrid Models | IMDB, Yelp, Amazon Reviews | 1.Covers sentiment evaluation, information categorization, QA, and natural language inference (NLI).  2. Organized into categories like RNNs, CNNs, Transformers, Capsule Networks, and Siamese Networks. Benchmarks provide insights into fine-performing models for precise NLP duties. | 1.Large-scale deep studying fashions require enormous assets for education and inference.  2.Models rely upon massive classified datasets, making them difficult to use in low-aid settings**.** | **[89]** |
| BERT  Seq2Seq | TREC QA,  Bing | **1.**In this article we are learning how to integrate deep learning techniques especially about transformer models such as BERT and GPT which has enhanced the MRC capabilities.  2.There are large data sets in this paper which generously helps in development and in training of MRC models efficiently. | **1.**Sometimes struggles in understanding of contexts, especially in longerpassages.  2.It requires large and high-quality datasets which may not be easily available. | **[90]** |
| DBNs and  LSTM | Stanford Natural Language Inference (SNLI) | 1.This paper extracts the hierarchical features of models very effectively.  2.DBNs is good for text extraction and classification. | Sometimes models struggle in sequencing data efficiently.  2.Requires time for training as data sets have large amount of data. | **[91]** |
| DCNN  CNN  HAN (Sentiment) | Sohu News | **1.** This paper shows how deep learning improves text classification by eliminating the text manually and enhancing accuracy.  2. CNN and RNN captures sequential dependencies and enhance interpretability. | 1.Models requires high power for performance.  2.Models are lacking in transparency in decision making topics. | **[92]** |
| Caps-Net-based | Amazon Reviews (User product reviews)  YouTube Music Ratings | 1.Caps-Net improve the CNN classification as it maintains relationship and avoid pooling operation.  2.It uses a special feature called gated sharing unit which filters out irrelevant features and improve efficiency. | It is more resource intensive and complex than RNN and CNN method.  2.It depends heavily on high quality datasets. | **[93]** |
| MLP and CNN | Sougo Lab's Sohu News | 1.Text classification is important in spam filtering, sentiment analysis, and information retrieval.  2.The application of CNN & RNN architectures enhances text classification accuracy. | 1.Limited to Chinese content only; other languages subject to varying penalties.  2.Pretrained embeddings required for best accuracy. | **[94]** |
| CRNN  HAN (Sentiment)  VDCNN | Yahoo Answers | 1.VDCNN is a improved version of CNN for text classification and improve performance over CNNs.  2.Directly works on Character instead of words for better working on different languages. | 1.As it uses advance version of CNN, hence require more power and time.  2.It struggles with tasks that requires long range dependencies. | **[95]** |
| BERT  MTL (Sentiment) | IMDB, MR, Amazon | 1.AMTL model helps in improving feature separation between shared and task specific spaces.  2.As data is shared, so it can be reused for more tasks. | 1.If data are not set in proper order than it can face overfitting issues. | **[96]** |
| RNNs | EHRs | 1.The study show us how we can deal with imbalanced class distribution with help of text classification.  2.Various models are used to classify texts. | 1.As it uses specific datasets, so finding may not apply on other domains.  2.It uses medical notes which compromises with personal information of others and raises privacy concerns. | **[97]** |
| Bi-LSTM (Sentiment)  DCLSTM | 8,292 news articles | 1.These models outperforms older models which helps in achieving higher accuracy.  2.It uses common features of CNkdmN, LSTM and MLP by combining them which helps in capturing better relationship in text data. | 1.Although 8,292 news articles were used for data set, but still, this is small number for training and may affect the results.  2.Multiple deep learning models used which obviously increases the load on machine and affect performance | **[98]** |

1. **Discussion & Conclusion**
2. **GAP ANALYSIS**

A Comprehensive Review of Deep Learning Architectures for Task Specific Analysis, provides a strong foundation by systematically reviewing deep learning models across key tasks such as reading comprehension, translation, summarization, question answering, generation, and reasoning. However, notable gaps exist, including limited coverage of emerging domains like multimodal learning and real-time AI, insufficient benchmarking of lightweight models for low-resource environments, and minimal focus on ethical challenges and bias mitigation strategies. The study primarily reviews well-known datasets and models but lacks detailed quantitative comparisons or real-world deployment case studies. Furthermore, it overlooks fast-growing trends like large language model (LLM) fine-tuning and retrieval-augmented generation. Addressing these gaps would significantly strengthen the study’s relevance, applicability, and contribution to current deep learning research.

1. **PROBLEM STATEMENT**

Despite significant advancements in deep learning, there remains a lack of comprehensive, task-specific analysis that systematically evaluates deep learning architectures across a wide range of applications. Current studies often focus on individual tasks or models without providing an integrated comparison, and they frequently overlook emerging challenges such as computational efficiency, data scarcity, ethical concerns, and real-world deployment issues. Additionally, the rapid evolution of lightweight models, fine-tuning methods, and domain-specific datasets demands updated reviews that address practical deployment needs and fairness. Therefore, there is a critical need for a structured and holistic review that not only summarizes existing architectures but also highlights their strengths, limitations, and future directions to guide researchers, practitioners, and industry professionals in selecting and optimizing deep learning models for task-specific applications.

1. **OBJECTIVES**

* To systematically review deep learning architectures used for various task-specific applications such as reading comprehension, translation, summarization, question answering, generation, and reasoning.
* To identify and analyze the datasets commonly employed across different deep learning tasks and examine the principles behind their design.
* To evaluate the strengths and limitations of popular deep learning models in terms of performance, scalability, interpretability, and real-world applicability.
* To highlight the challenges faced in deep learning tasks, including issues of computational demands, data scarcity, overfitting, generalization, and ethical considerations.
* To provide comparative insights that assist researchers and practitioners in selecting, adapting, and improving models based on specific task requirements.
* To suggest future directions by identifying gaps in current research and proposing areas for improvement, including emerging trends like lightweight models, explainable AI, and retrieval-augmented generation techniques.

**CHAPTER 3: METHODOLOGY**

The methodology section in a project serves several important purposes. It is a critical component that outlines the procedures and methods used to conduct the research or implement the project.

3.1 **Overall architecture /Flow chart**:

1. Google Scholar (Data Collection Stage)

What it means:

The research process began by collecting academic papers and research articles from Google Scholar, a trusted source for scholarly information.

Purpose:

To gather updated, relevant, and credible research about deep learning architectures and their use in different tasks.

2. Deep Learning (Central Research Focus)

What it means:

After gathering information, the main focus was placed on Deep Learning models — analyzing how different architectures work for various specific tasks.

Purpose:

To systematically study and understand how deep learning is applied for task-specific problem solving.

3. Task-Specific Breakdown

The research further divides Deep Learning applications into several key tasks:

Task Explanation

Reading Comprehension Studying how models like BERT and GPT understand and answer questions based on text passages.

Translation Analyzing how deep learning models convert text from one language to another (e.g., using Transformer models).

Summarization Exploring how models generate short summaries of large documents (using models like T5, BART).

Question and Answer Session Investigating how deep models directly answer user queries from datasets or documents (like SQuAD datasets).

Generation Studying how models create new content such as text, audio, or synthetic data (e.g., GPT, VAEs, GANs).

Reasoning Understanding how models simulate logical thinking or problem solving (e.g., GPT-4, BERT for reasoning).

Text Classification Although not initially emphasized, this part involves sorting or tagging texts into categories (e.g., spam detection, sentiment analysis).

Summary of Flow

First, papers were collected (Google Scholar).

Second, research was focused around Deep Learning models.

Third, deep learning applications were divided into different specialized tasks, and each task was analyzed separately to study the models, datasets, advantages, challenges, and future directions.

A diagram of a company

AI-generated content may be incorrect.

Figure 1. Research Design and methodology

**Chapter 4**

**Implementation**

The project was implemented using a systematic review methodology. The primary goal was to collect, study, and analyze various deep learning architectures applied to different specific tasks. First, academic papers and research articles were collected from Google Scholar using structured keyword searches. Keywords such as "Deep Learning Architectures", "Task-Specific Deep Learning", "Text Summarization", "Translation Models", and "Question Answering Systems" were used.

Once the data was collected, papers were filtered based on relevance, recent publication dates, and task-specific focus. The selected papers were then categorized based on six main tasks: Reading Comprehension, Translation, Summarization, Question and Answer Session, Generation, and Reasoning.

For each task, deep learning models (e.g., BERT, GPT-2, T5, Transformer) and datasets (e.g., SQuAD, CNN/Daily Mail, WMT) were identified. Their performance, strengths, limitations, and applicability were reviewed in detail. Finally, challenges in deep learning applications were documented, and gaps in the existing research were analyzed to provide future recommendations.

Since this research was a systematic review paper (theoretical analysis), no coding or model training was directly implemented. Instead, the study analyzed the underlying algorithms of popular deep learning architectures:

Transformer Algorithm: Based on self-attention mechanism to capture dependencies in sequences.

BERT: Based on bidirectional transformers with masked language modeling and next sentence prediction.

GPT: Based on autoregressive transformer models to generate text word-by-word.

Discussion of Any Challenges Faced During Implementation and Their Solutions

Challenges Faced:

Large Volume of Research Papers:

It was difficult to filter only the most relevant and high-quality studies from a large volume of publications.

Rapid Evolution of Deep Learning:

New models and architectures (like GPT-4, PaLM) are being released continuously, making it hard to keep the review updated.

Limited Access to Some Datasets or Papers:

Some advanced datasets or paywalled papers could not be accessed directly.

Difficulty in Comparative Analysis:

Different studies used different evaluation metrics, making it hard to directly compare models.

Solutions Applied:

Systematic Keyword Filtering and Recent Literature Selection:

Focused only on papers published within the last 5 years and matched task-specific keywords.

Scope Limitation:

Restricted the review to the most commonly cited models and datasets to ensure quality over quantity.

Use of Open Access Resources:

Prioritized papers and datasets that were freely available or open source.

Task-wise Comparative Tables:

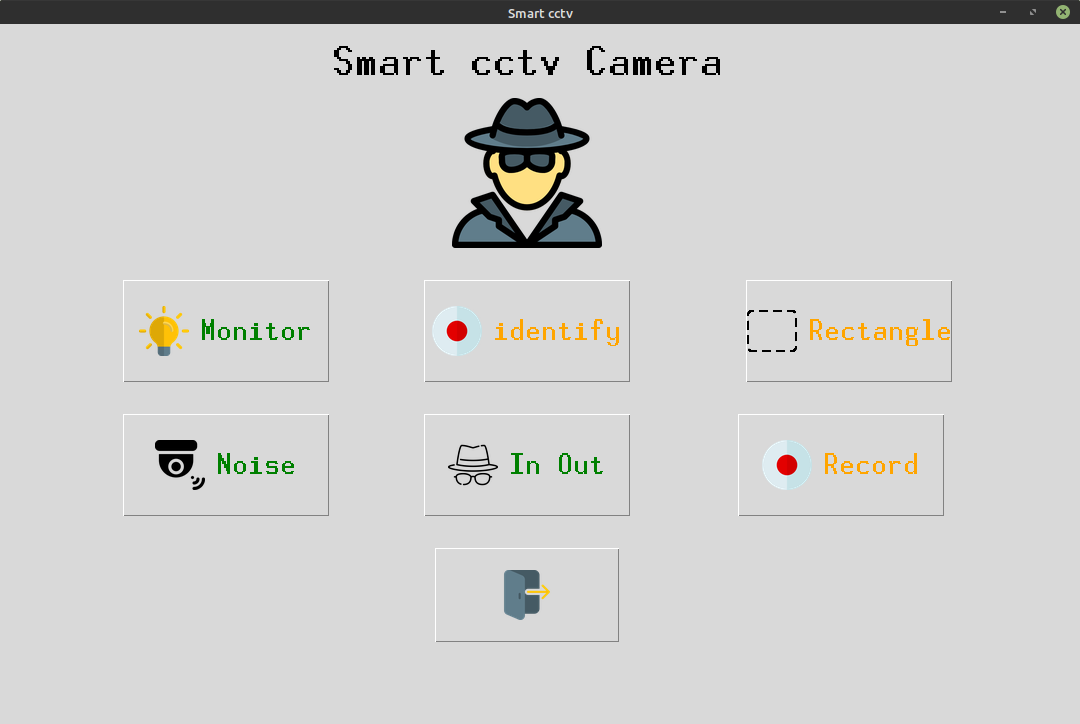
Created comparative tables for each task, summarizing models, datasets, key takeaways, and limitations uniformly for easy comparison.

**Chapter 5**

**RESULTS AND DISCUSSIONS**

This study systematically reviews deep learning tasks, including reading comprehension, translation, text generation, question answering, reasoning, summarization, and classification. It highlights the significant advancements made in these areas, alongside the persistent challenges that remain. Over time, numerous generalized and specialized models and diverse datasets have been developed and utilized to address these specific tasks effectively. However, the variability in design and methodology across these models and datasets demonstrates the complexity of developing solutions that can be universally applied. A snapshot of the observed models and datasets is provided in Figure S1-S8. The models under review, ranging from transformers to RNNs, all have something to offer in terms of strength, with some being more scalable, and yet others having greater context-sensitive task accuracy. Core datasets, which have been instrumental in model training, are central to determining the outcome of deep learning models. They form the basis for measuring model performance, directing researchers toward solving specific challenges inherent in various domains. Yet, dataset design itself brings with it limitations—bias, domain specificity, and generalization issues—that must be given careful thought when choosing or designing datasets for task-specific use. Assessment of models and their respective datasets point to the general trend of incremental performance improvement over multiple tasks, with transformer-based models, especially BERT, GPT, and T5, dominating tasks such as reading comprehension, question answering, and summarization. These models tend to utilize large-scale, high-variance datasets like SQuAD, GLUE, and CNN/Daily Mail, offering an abundance of training data but also revealing some limitations in terms of domain transferability and bias. For generation and translation tasks, GPT-based models excel but continue to struggle with nuances of languages and produce contextually consistent results in lengthy text formats. Reasoning tasks, though making advancements through models such as T5 and GPT-3, continue to need improvements in terms of logical reasoning strength and common-sense inference. Summarization and classification work perform well on extractive and abstractive tasks. Despite the strengths, concerns exist for factual accuracy and coherence retention in automatic summaries, particularly in domain-specialized cases. Conversely, many state-of-the-art models lack proper handling of aspects like data bias, model bias, heavy computationally demanding work, and lesser explainability, which negatively contribute to their utilization in practical use cases where there is a demand for transparency as well as economic resource optimization.

Although task-specific models have demonstrated significant accuracy and efficiency improvements, the review calls out a number of key areas for further investigation. These include improving generalization across languages and domains, dealing with ethical issues such as bias and fairness, and enhancing interpretability to provide transparency in decision-making. Moreover, merging multimodal datasets and models, applying transfer learning, and making models adaptable to low-resource languages are areas that show promise for further developing deep learning models in these tasks.



**Chapter 6**

**CONCLUSION**

In recent decades, there have been a vast number of reasons contributing to the unexpectedly growing crimes. Urbanization, rapid economic liberalization, growing large-scale political turmoil, fierce conflicts, and inadequate and inappropriate policies can be listed as the basis of crime in urban areas. Moreover, crime rate has significantly increased due to current pandemic, and has made things worse for the security officials of all countries.

The closed-circuit television (CCTV) is one of the devices used to monitor the secured area for any intruders. But the traditional CCTVs have their own set of flaws, which make them less effective for securing the area. This project tries to implement Machine Learning algorithms to enhance the working of traditional CCTVs, by adding functionalities:

1. **Monitor** – monitors the area under surveillance.
2. **Identify** – Identifies the family members.
3. **Noise** **Detection**– Finds any motion in the frame.
4. **In** **Out** **Detection** – Finds who enters and exits.

This will help make a stronger system for security concerns, and will make the users feel more secure when they are not at home. It will not only mitigate the risks of crime occurrence, but also capture anything and everything that can be considered as a proof against the criminal , provided the crime takes place.

**FUTURE WORK**

Theproject is working well for all the features including- monitoring, facial detection, noise recognition and in-out movement detection. The future work can include enhancing the model by using a better face recognition approach like Dlib, as haarcascade classifier is not very accurate at all times. The project is scaled for a very limited user base, so it can only detect a little motion and only a couple entries and exits simultaneously. It can be scaled up for more number of users and thus can be used in other domains like bank security management, school security, etc. Better technology for night vision can be implemented as it is not appropriate for darker regions.

**Publication Details with proof**

**Add a relevant proof here**

**REFERENCES**

1. Kim, J., & Kim, H. J. (2019). A Smart CCTV System Based on Deep Learning for Real-time Human Detection in a Construction Site. Automation in Construction, 106, 102873.
2. Gupta, A., Kumar, S., & Chakraborty, S. (2020). IoT Based Smart Surveillance System Using CCTV Camera and Raspberry Pi. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
3. Zhou, S., & Huang, R. (2017). Smart Video Surveillance System Based on IoT and Cloud Computing. Future Generation Computer Systems, 76, 319-326.
4. Sankaranarayanan, S., Al-Maadeed, S., & Trivedi, M. M. (2019). Smart CCTV Video Analysis and Anomaly Detection in Traffic and Transportation: A Survey. IEEE Transactions on Intelligent Transportation Systems, 21(1), 256-269.