Libyan Academy of Postgraduate Studies Tripoli, Libya
School of Applied Sciences and Engineering
Department of Electrical and Computer Engineering
Division of Information Technology (PHD)

A Superior Arabic Text Categorization Deep Model (SATCDM)

By
M. ALHAWARAT , (Member, IEEE), AND AHMAD O. ASEERI *IEEE Access 2020.*

Read & Summarized by Amal Omar Saad & Abdulmenam Alahreah
Subject: Natural Language Processing(NLP)
Lecturer: Dr. Abdulbaset Goweder

Outlines

- Introduction
- Related Work
- Contribution
- Convolutional Neural Networks (CNNs)
- Data Preparation
- SATCDM Model.
- Methodology and Experiments Setup
- Results and Discussion
- Conclusion
- Critique

Introduction
“Problem”

- The number of Arabic increases significantly each day. Need to a good classification.
- The nature of the language itself having rich dialects and enormous numbers of synonyms**.**
- The lack of Arabic resources e.g, inaccurate stemming algorithms.

Related Work

- There have been many studies that applied traditional ML algorithms for document classification including Support Vector Machine (SVM), Naïve Bayes classier (NB), K-Nearest Neighbor (KNN), Decision Trees (DT) and Rocchio classier.
- Some of them applied or developed stemmer to traditional ML classifiers (e.g, P-stemmer)
- Few studies have used deep learning techniques for Arabic document classifications, including social media text.
- Recent researches, use Convolutional Neural Network (CNN) with stemming algorithms, Frequency-Inverse Document Frequency (TD-IDF).

Related Work

Ref	Model	Compression	Results
24	Built and investigated the word embedding model for sentiment classification for the Arabic language	Other word embedding	Outperformed
25	Proposed a three-stage algorithm to classify Arabic documents using deep belief networks and Markov clustering.	NB, KNN, and SVM algorithms	Outperformed
26	Classified Arabic documents based on document embeddings (doc2vec)	ML	Better results
20	Constructed a dataset of 237,000 Arabic news articles and then applied traditional ML algorithms such as SVM, NB, and Random Forest developed a new stemming algorithm called p-stemmer	SVM, NB, and Random Forest	Best result was using SVM
28	Convolutional Neural Network (CNN) with word Embedding by built a word embedding	Methods that rely on linear SVM algorithm and SVM-BOW	Outperforms other existing methods
29	Compared the performance of SVM, NB, and KNN	Deep Neural Networks	The best two are deep neural networks and NB classier
30	Use Convolutional Neural Networks to classify Arabic documents. 1. use stemming algorithms to minimize the number of features. 2. TD-IDF to weigh and select important features as input to the CNN.	ML algorithms	Better results of proposed model
31.32	Use three Convolutional layers	Different deep learning models	CNN has better results

AIM and CONTRIBUTION

- To helping researchers in the field of ANLP to classify Arabic text documents **more correctly into predefined classes.**

- To achieve better results using the latest deep learning technology and algorithms, **including:**
  - **Convolutional Neural Network (CNN)**
  - **word embedding (word2vec).**

PROPOSED SOLUTION

- Since then, deep learning has made amazing strides in **computer vision** and **voice.**
- A Multi-Kernel CNN model for classifying Arabic news documents enriched with **word2vec** (n-gram word embedding**)** is proposed.
- Proposed model called (**Superior Arabic Text Categorization Deep Model (SATCDM)**) is presented.

METHODOLOGY

- Utilizes an efficient multi-kernel CNN architecture.
- Uses a **skip-gram word embedding** language model enriched with sub-word.
- Uses free 15 datasets representing Arabic News text documents that comes in **Modern Standard Arabic Format(MSA)** format.
  - **Two Small-Medium size datasets:**
    - with original text.
    - with stemmed text.
  - **One Large-size datasets with original text.**
- Compare the results with traditional ML techniques that are usually used in performing the Arabic text classification.

BACKGROUNG

Concepts of Convolutional Neural Networks (CNN)

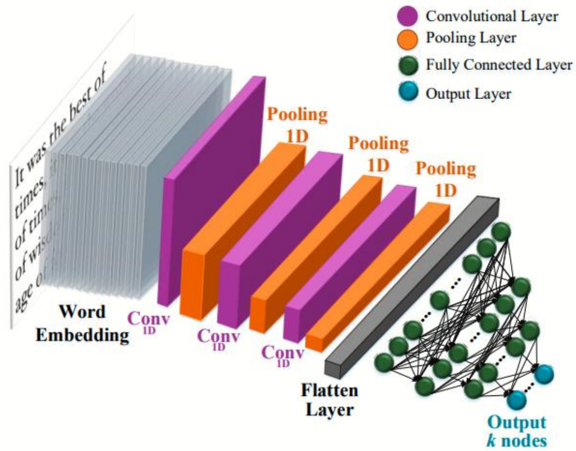


Figure (1):CNN
Source : Text Classification Algorithms: A Survey

Cont .. BACKGROUNG Concepts of (CNN)

A generic CNN architecture for Text Categorization

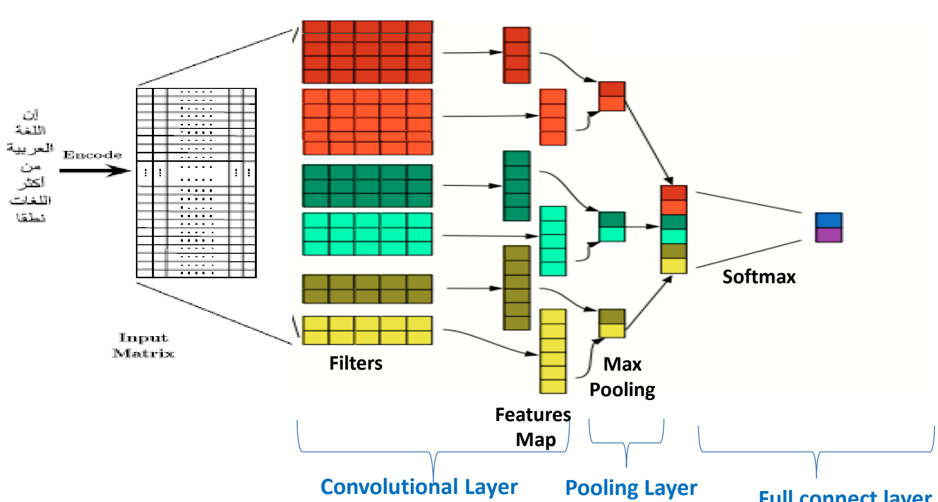


Figure (2):. A generic CNN architecture for Text Categorization

Cont ... Proposed model

SATCDM Architecture

Word Embedding (1000)

Three convolution layers use 128 filters and kernel sizes 2, 3, and 5, respectively

Batch-Normalization layer

Dropout layer

ReLU activation function

Global-Max-Pooling

Dense Layer

SoftMax Activation Function

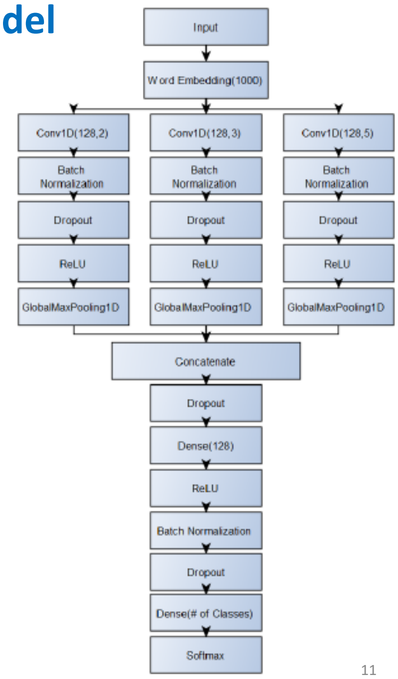


Figure (3):. . SATCDM architecture

Cont ... METHODOLOGY

Word Embedding

Word embedding :
Is a text representation currently used to represent text terms **as real-valued vectors** in vector space that represent both **syntactic** and **semantic** traits of text.

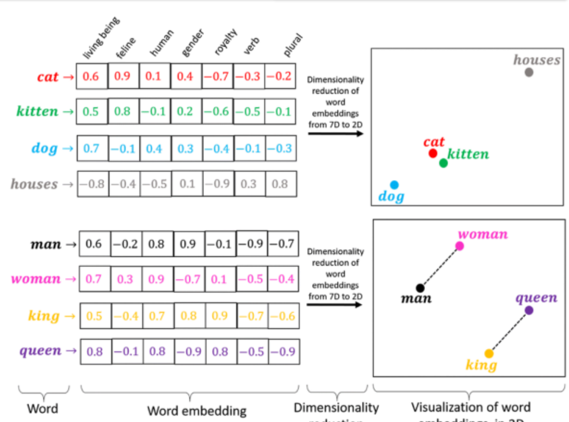


Figure (4): Word Embedding
Source: https://medium.com/@hari4om/word-embedding-d816f643140

Cont … **METHODOLOGY**

Word Embedding

- In this study, the **Arabic Wikipedia fast-text model** (***skip gram model***) is used
  - Trained on **Wikipedia using fastText**. These vectors in dimension 300 were obtained using the skip-gram model
  - It consists of around **610k unique tokens** and is available for many languages, including Arabic.

Cont … **METHODOLOGY**

How does Convolution Layer Work?

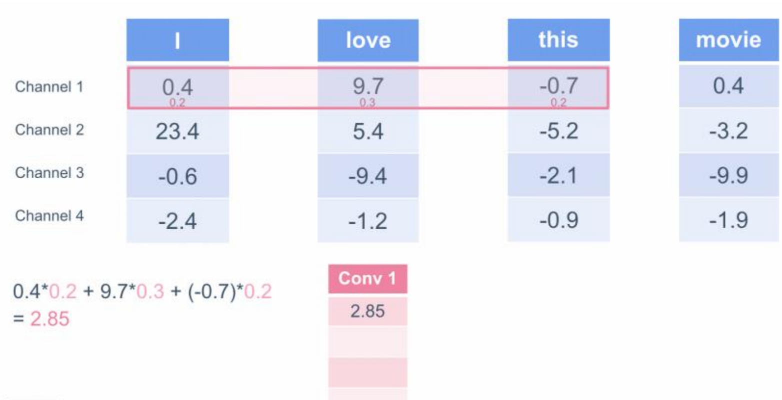


Figure (6): Convolution Layer Work

Cont … **METHODOLOGY**

Batch-Normalization layer

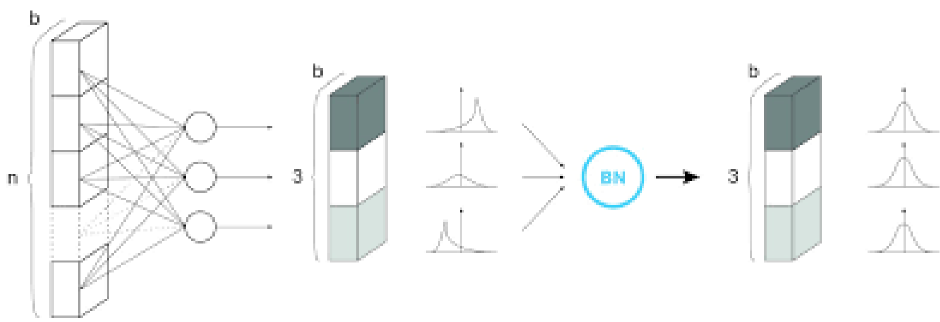


Figure (9): Batch-Normalization layer
Sources: https://towardsdatascience.com/batch-normalization-in-3-levels-of-understanding-14c2da90a338

Cont … **METHODOLOGY**

Dropout Layer

- To prevent over fitting problem.

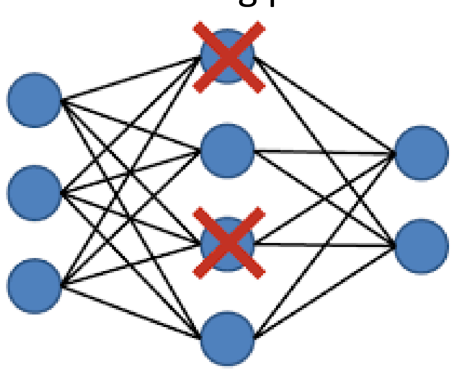


Figure (10):. Dropout Layer
Sources:

Cont … **METHODOLOGY**

ReLU Activation Function

Rectified Linear Units (ReLUs) is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input.

The function is understood as: f(x)=max(0,x)

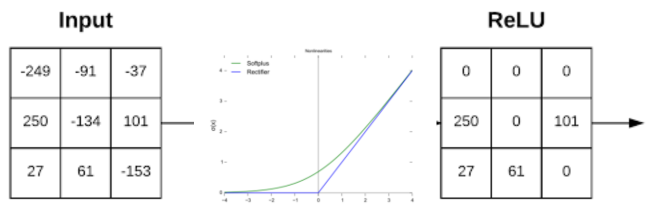


Figure (11): ReLU Activation Function
Sources: https://deepai.org/machine-learning-glossary-and-terms/rectified-linear-units

Cont … **METHODOLOGY**

Global-Max-Pooling

Using pooling layers to reduce the number of parameters and computation cost, making it ideal for controlling overtting.

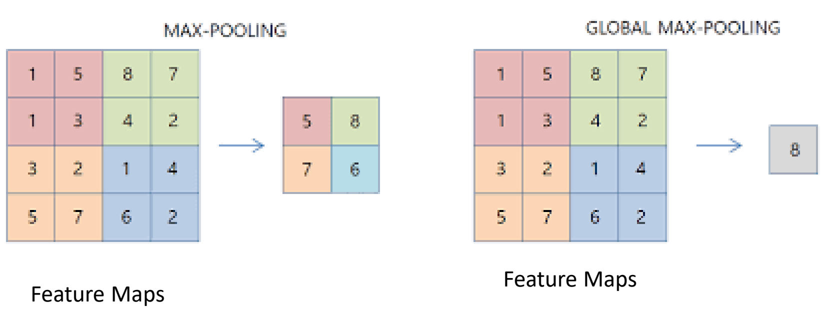


Figure (12):. Max-Pooling and Global Max-Pooling

Cont … **METHODOLOGY**

Dense Layer

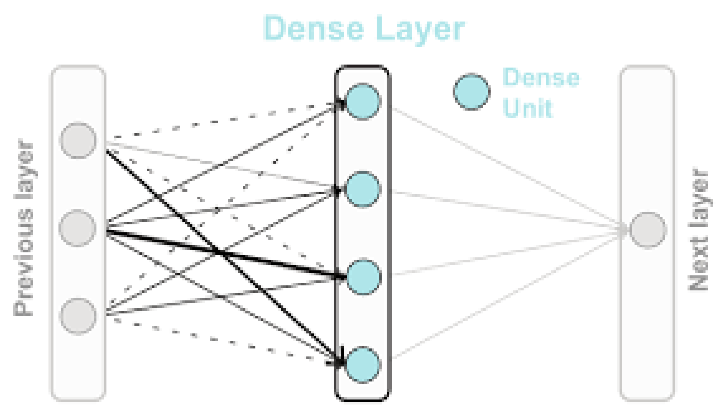
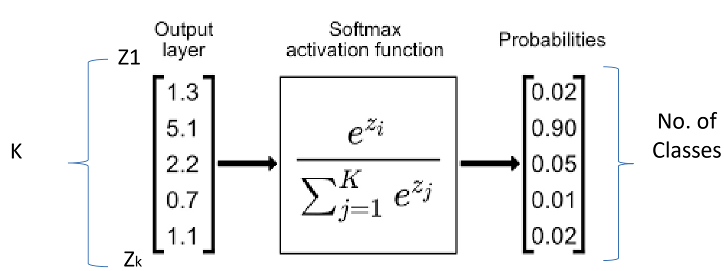


Figure (13):.Dense Layer

Cont … **METHODOLOGY**

Softmax Activation Function

- **Multiclass Classification Problem**



e=2.718

Figure (14): Softmax Equation

Cont ...

Experiments Setup

- There are mainly two experiments:
  - Traditional ML algorithms (SVM, SGD classifier, and NB).
  - SATCDM model.
- Experimental Setup:
  - The python programming language is used where Keras and TensorFlow.
  - Each of these experiments is applied to all 15 datasets.
  - Platforms:
    - For traditional machine learning a PC with i7 core and 32GB RAM equipped with a GTX-1070 GPU with 8GB RAM is used.
    - Whereas for Deep learning model we use the free K80 online GPU with 24GB RAM that is available through **Google Colab** environment.

Cont ...

Experiments Setup

- Avoiding Overfitting and Selection Bias:
- Dropout and Batch Normalization.
- Cross-validation
  - 5-fold cross-validation technique
  - Dividing the dataset into three parts: train, validation, and test sets:
    - **Traditional ML:** 20% for testing, 80% for training (divided into: 64% for training and 16% for validation for **each fold**).
    - **Deep Learning:** 10% for testing, 90% for training (divided into: 72% for training and 18% for validation for **each fold**).

Cont …

Experiments Setup

- Not all features are used during training the model:
  - for datasets with less than 400,000 unique tokens, then 40% of the most frequent tokens are used
  - if the dataset contains more than 400,000 unique tokens then only 25% is used.
  - Finally, out of vocabulary (OOV), tokens are set to zeros.
  - the most 1000 frequent terms are used;

These assumptions are decided experimentally

DATA PREPARATION

Datasets are classified into three categories:

- **Small-Medium size datasets with original text:**

TABLE 1. Details of the small-medium size datasets with original text.

Dataset	Nº Classes-Docs.	Nº Terms-Unique Terms	Doc. Avg. Length
Abuaiadah(V1) [12]	9 - 2,700	878,726 - 96,859	325
Aljazeera [37]	5 - 1,500	388,653 - 50,099	259
Alwatan [38]	6 - 20,291	9,876,786 - 261,909	487
Alkhaleej [39]	4 - 5,690	2,472,763 - 122,162	435
OSAC [40]	10 - 22,429	18,183,511 - 677,972	811
BBC [40]	7 - 4,763	1,794,123 - 88,953	377
CNN [40]	6 - 5070	2,166,109 - 105,047	427

- **Small-Medium size datasets with stemmed text:**

TABLE 2. Details of the small-medium size datasets with stemmed text.

Dataset	Nº Classes-Docs.	Nº Terms-Unique Terms	Doc. Avg. Length
Abuaiadah(V2) [12]	9 - 2,700	600,627 - 89,757	325
Abuaiadah(V3) [12]	9 - 2,700	600,552 - 42,571	222
Abuaiadah(V4) [12]	9 - 2,700	600,477 - 30,488	222
Abuaiadah(V5) [12]	9 - 2,700	600,602 - 13,803	222
NADA [44]	10 - 7,310	3,248,653 - 152,050	444

- V1: original dataset type.
- V2: no stop words.
- V3: **V2 where Light10 stemming algorithm** is applied.
- V4: V2 where Chen's stemming algorithm is applied.
- V5: V2 where Khoja's stemming algorithm is applied.
- Feature selection algorithm is used to reduce the dimensionality of the dataset from 22k text documents to around 7k

DATA PREPARATION

– Large-size datasets with original text:

- dataset comprises of a large number of news documents wit ha total of around 195 thousand

TABLE 3. Details of the large-size datasets for SANAD [45].

Dataset	Nº Classes-Docs.	Nº Terms-Unique Terms	Doc. Avg. Length
Arabiya	6 - 71,246	16,817,633 - 429,597	236
Khaleej	7 - 45,500	16,801,740 - 689,155	369
Akhbarona	7 - 78,428	19,965,764 - 772,872	255
All Combined	7 - 195,174	53,585,137 - 1,295,948	275

– Data Preprocessing :

- No preprocessing for the input text documents .
- Some of the datasets have already been preprocessed by their creators.

Results and Discussion

Small-medium Size Datasets With Original Text

- 1.5k - 22k text documents with 50k-678k terms.

TABLE 4. Accuracy for small-medium size datasets with original text.

Dataset	SATCDM	SVM	SGD	NB
Abuaiadah(V1)	98.93%	97.15%	96.81%	95.93%
Aljazeera	98.77%	96.33%	96.60%	96.07%
Alwatan	98.11%	94.89%	92.01%	87.15%
Alkhaleej	97.58%	95.45%	94.06%	83.41%
OSAC	99.90%	99.32%	96.18%	92.68%
BBC	99.76%	90.30%	85.95%	62.69%
CNN	97.63%	94.38%	92.44%	73.39%

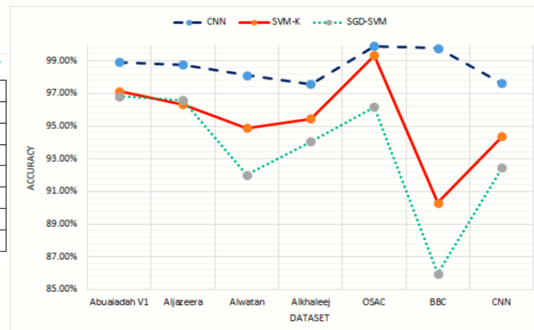


Figure (15): **Accuracy for datasets with original text.**

Small-Medium Size Datasets with Stemmed Text

- 2.7k to 7.3k documents terms 14k-152k.

TABLE 6. Accuracy for small-medium size datasets with stemmed text.

Dataset	SATCDM	SVM	SGD	NB
Abuaiadah(V2)	98.89%	97.15%	97.04%	96.44%
Abuaiadah(V3)	98.52%	97.11%	96.85%	96.30%
Abuaiadah(V4)	98.64%	97.07%	96.89%	96.04%
Abuaiadah(V5)	98.68%	96.67%	96.44%	95.22%
NADA	99.75%	99.40%	98.85%	82.43%

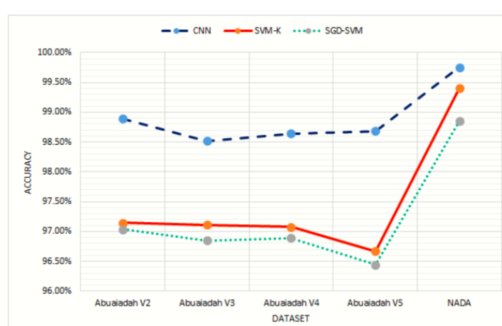


Figure (16): **curacy for small-medium size with stemmed texts datasets.**

Large Size Datasets

- 45.5k to 78.5k documents and 429k-772k terms.

TABLE 8. Accuracy for large-size datasets with original text.

Dataset	SATCDM	SVM	SGD	NB
Arabiya	99.49%	98.36%	95.62%	85.79%
Khaleej	99.57%	98.16%	96.87%	95.80%
Akhbarona	98.44%	94.88%	91.27%	87.45%
All-Combined	98.80%	96.02%	93.30%	90.75%

Impact of Stemming

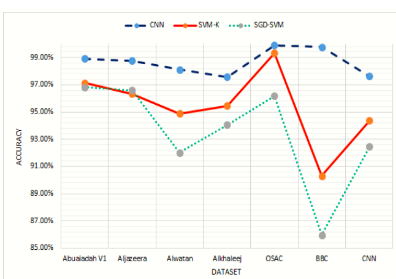


FIGURE 3. Accuracy for datasets with original text.

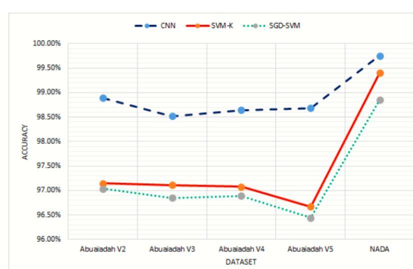
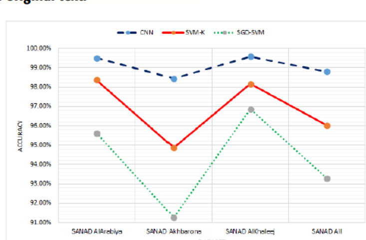


FIGURE 4. Accuracy for small-medium size with stemmed texts datasets.

Large data set



GURE 5. Accuracy for SANAD datasets.

Conclusion

- The presented model achieves superior accuracy results compared to similar studies
- is suitable for any Arabic text documents regardless of normalization, preprocessing, stemming algorithms, or methods.
- More research opportunities are available for more accuracy gain using other methods and models, including recurrent models such as Long Short-Term Memory (LSTM) and Gated recurrent units (GRUs) in addition to attention models.

Critique The Paper

- 99.9 % ?
- Although the superior results the Citation for this paper is still low.
- A Little description about the proposed model.

References

[1] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Trans. Asian Lang. Inf. Process.*, vol. 8, no. 4, pp. 14:1–14:22, Dec. 2009.

[2] M. Saad and W. Ashour, "Arabic morphological tools for text mining," in *Proc. 6th Int. Symp. Electr. Electron. Eng. Comput. Sci.* Lefke, Cyprus: European University Lefke, 2010, pp. 112–117.

[3] A. Schmidt and M. Wiegand, "A survey on hate speech detection using natural language processing," in *Proc. 5th Int. Workshop Natural Lang. Process. Social Media*, 2017, pp. 1–10.

[4] X. Xue-Feng and Z. Guo-Dong, "A survey on deep learning for natural language processing," *Acta Autom. Sinica*, vol. 42, no. 10, pp. 1445–1465, 2016.

[5] I. Hmeidi, M. Al-Ayyoub, N. A. Abdulla, A. A. Almodawar, R. Abooraig, and N. A. Mahyoub, "Automatic Arabic text categorization: A comprehensive comparative study," *J. Inf. Sci.*, vol. 41, no. 1, pp. 114–124, Feb. 2015.

[6] A. H. Mohammad, T. Alwada'n, and O. Al-Momani, "Arabic text categorization using support vector machine, Naïve Bayes and neural network," *GSTF J. Comput.*, vol. 5, no. 1, p. 108, 2016.

[7] I. Hmeidi, B. Hawashin, and E. El-Qawasmeh, "Performance of KNN and SVM classifiers on full word Arabic articles," *Adv. Eng. Inform.*, vol. 22, no. 1, pp. 106–111, Jan. 2008.

[8] M. A. H. Madhfar and M. A. H. Al-Hagery, "Arabic text classification: A comparative approach using a big dataset," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Apr. 2019, pp. 1–5.

[9] M. S. Khorsheed and A. O. Al-Thubaity, "Comparative evaluation of text classification techniques using a large diverse Arabic dataset," *Lang Resour. Eval.*, vol. 47, no. 2, pp. 513–538, Jun. 2013.

[10] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882

[11] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 135–146, Dec. 2017.

[12] D. Abuaiadah, J. E. Sana, and W. Abusalah, "Article: On the impact of dataset characteristics on Arabic document classification," *Int. J. Comput. Appl.*, vol. 101, no. 7, pp. 31–38, Sep. 2014.

[13] H. Sawaf, J. Zaplo, and H. Ney, "Statistical classification methods for Arabic news articles," in *Proc. Natural Lang. Process. (ACL)*. Toulouse, France: Citeseer, 2010.

[14] A. Abu-Errub, "Arabic text classification algorithm using TFIDF and chi square measurements," *Int. J. Comput. Appl.*, vol. 93, no. 6, pp. 40–45, May 2014.

[15] S. Al-Harbi, A. Almuhareb, A. Al-Thubaity, M. S. Khorsheed, and A. Al-Rajeh, "Automatic arabic text classification," in *Proc. 9th Journees Internationales d'Analys e statistique des Donnees Textuelles (JADT)*, vol. 9, 2008.

[16] A. M. El-Halees, "Arabic text classification using maximum entropy," *Arabic Text Classification Using Maximum Entropy*, vol. 15, no. 1, pp. 157–167, 2007.

[17] M. El Kourdi, A. Bensaid, and T.-E. Rachidi, "Automatic Arabic document categorization based on the Naïve Bayes algorithm," in *Proc. Workshop Comput. Approaches Arabic Script-Based Lang.*, 2004, pp. 51–58.

[18] T. F. Gharib, M. B. Habib, and Z. T. Fayed, "Arabic text classification using support vector machines," *IJ Comput. Appl.*, vol. 16, no. 4, pp. 192–199, 2009.

[19] G. Kanaan, R. Al-Shalabi, S. Ghwanmeh, and H. Al-Ma'adeed, "A comparison of text-classification techniques applied to Arabic text," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 9, pp. 1836–1844, Sep. 2009.

[20] T. Kanan and E. A. Fox, "Automated Arabic text classification with P-Stemmer, machine learning, and a tailored news article taxonomy," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 11, pp. 2667–2683, Nov. 2016.

[21] A. H. Mohammad, O. Al-Momani, and T. Alwada'n, "Arabic text categorization using k-nearest neighbour, decision trees (C4. 5) and Rocchio classifier: A comparative study," *Int. J. Current Eng. Technol.*, vol. 6, no. 2, pp. 477–482, 2016.

[22] R. Alshammari, "Arabic text categorization using machine learning approaches," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 1–5, 2018.

[23] M. K. Saad, "The impact of text preprocessing and term weighting on Arabic text classification," M.S. thesis, Islamic Univ.-Gaza, Gaza, Palestine, 2010.

[24] A. A. Altowayan and L. Tao, "Word embeddings for Arabic sentiment analysis," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 3820–3825.

[25] V. Jindal, "A personalized Markov clustering and deep learning approach for Arabic text categorization," in *Proc. ACL Student Res. Workshop*, 2016, pp. 145–151.

[26] A. El Mahdaouy, E. Gaussier, and S. O. El Alaoui, "Arabic text classification based on word and document embeddings," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.* Cham, Switzerland: Springer, 2016, pp. 32–41.

[27] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of Arabic word embedding models for use in Arabic NLP," *Procedia Comput. Sci.*, vol. 117, pp. 256–265, Jan. 2017.

[28] A. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, and P. Duan, "Word embeddings and convolutional neural network for Arabic sentiment classification," in *Proc. 26th Int. Conf. Comput. Linguistics (COLING)*, 2016, pp. 2418–2427.

[29] M. Sayed, R. Salem, and A. E. Khedr, "Accuracy evaluation of Arabic

[30] M. Biniz, S. Boukil, F. El Adnani, L. Cherrat, and A. E. El Moutaouakkil, "Arabic text classification using deep learning techniques," *Int. J. Grid Distrib. Comput.*, vol. 11, pp. 103–114, 09 2018.

[31] A. Elnagar, O. Einea, and R. Al-Debsi, "Automatic text tagging of Arabic news articles using ensemble deep learning models," in *Proc. 3rd Int. Conf. Natural Lang. Speech Process.*, 2019, pp. 59–66.

[32] A. Elnagar, R. Al-Debsi, and O. Einea, "Arabic text classification using deep learning models," *Inf. Process. Manage.*, vol. 57, no. 1, 2020, Art. no. 102121.

[33] M. Al-Ayyoub, A. Nuseir, K. Alsmearat, Y. Jararweh, and B. Gupta, "Deep learning for Arabic NLP: A survey," *J. Comput. Sci.*, vol. 26, pp. 522–531, May 2018.

[34] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[35] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 25. San Francisco, CA, USA: Determination Press, 2015.

[36] M. Sarigül, B. Ozyildirim, and M. Avci, "Differential convolutional neural network," *Neural Netw.*, vol. 116, pp. 279–287, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608019301315

[35] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 25. San Francisco, CA, USA: Determination Press, 2015.

[36] M. Sarigül, B. Ozyildirim, and M. Avci, "Differential convolutional neural network," *Neural Netw.*, vol. 116, pp. 279–287, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608019301315

[37] D. Said, N. M. Wanas, N. M. Darwish, and N. Hegazy, "A study of text preprocessing tools for Arabic text categorization," in *Proc. 2nd Int. Conf. Arabic Lang.*, 2009, pp. 230–236.

[38] M. Abbas, K. Smaïli, and D. Berkani, "Evaluation of topic identification methods on Arabic corpora," *J. Digit. Inf. Manage.*, vol. 9, no. 5, pp. 185–192, 2011.

[39] M. Abbas and K. Smaili, "Comparison of topic identification methods for Arabic language," in *Proc. Int. Conf. Recent Adv. Natural Lang. Process. (RANLP)*, 2005, pp. 14–17.

[40] M. K. Saad and W. Ashour, "OSAC: Open source Arabic corpora," in *Proc. 6th ArchEng Int. Symp., 6th Int. Symp. Elect. Electron. Eng. Comput. Sci. (EEECS)*, vol. 10, 2010, pp. 118–123.

[41] L. S. Larkey, L. Ballesteros, and M. E. Connell, *Light Stemming for Arabic Information Retrieval*. Dordrecht, The Netherlands: Springer, 2007, pp. 221–243.

[42] A. Chen and F. C. Gey, "Building an Arabic stemmer for information retrieval," in *Proc. TREC*, 2002, pp. 631–639.

[43] S. Khoja and R. Garside, "Stemming Arabic text," M.S. thesis, Dept. Comput., Lancaster Univ., Lancaster, U.K., 1999.

[44] N. Alalyani and S. L. Marie-Sainte, "NADA: New Arabic dataset for text classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 1–7, 2018.

[45] O. Einea, A. Elnagar, and R. Al Debsi, "SANAD: Single-label Arabic news articles dataset for automatic text categorization," *Data Brief*, vol. 25, Aug. 2019, Art. no. 104076.

[46] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: https://arxiv.org/abs/1301.3781

[47] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[48] G. Khazal *et al.*, "Arabic word embedding method for NLP," in *Proc. Inf. Technol. Math. Modeling (ITMM)*, vol. 6, 2019, pp. 99–105.

[49] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

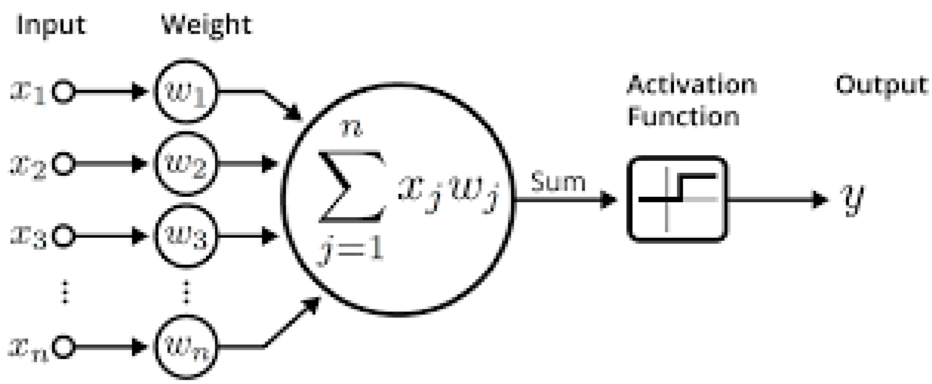
[51] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[53] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. Eur. Conf. Mach. Learn.* Berlin, Germany: Springer, 1998, pp. 137–142.

Thank you for your attention

Concepts of Neural Network(NN)



Concepts of Convolutional Neural Network (CNN)