REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf viii, 1, 4, 6, 8

[2] J. Pfeiffer, I. Vulić, I. Gurevych, and S. Ruder, "MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 7654–7673. [Online]. Available: https://aclanthology.org/2020.emnlp-main.617 viii, 13, 20, 37, 38, 47

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423 1, 2, 7, 10

[4] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019. [Online]. Available: http://arxiv.org/abs/1907.11692 1, 9

[5] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *CoRR*, vol. abs/1910.10683, 2019. [Online]. Available: http://arxiv.org/abs/1910.10683 1

[6] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018, pp. 353–355. [Online]. Available: https://aclanthology.org/W18-5446 1

[7] A. Conneau, R. Rinott, G. Lample, A. Williams, S. Bowman, H. Schwenk, and V. Stoyanov, "XNLI: Evaluating cross-lingual sentence representations," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2475–2485. [Online]. Available: https://aclanthology.org/D18-1269 1, 35

[8] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury, "The state and fate of linguistic diversity and inclusion in the nlp world," *arXiv preprint arXiv:2004.09095*, 2020. 1, 43

[9] Y. K. Lal, R. Singh, H. Trivedi, Q. Cao, A. Balasubramanian, and N. Balasubramanian, "IrEne-viz: Visualizing energy consumption of transformer models," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 251–258. [Online]. Available: https://aclanthology.org/2021.emnlp-demo.29 1

[10] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?" in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 4996–5001. [Online]. Available: https://aclanthology.org/P19-1493 1, 2, 12

[11] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 8440–8451. [Online]. Available: https://aclanthology.org/2020.acl-main.747 2, 9, 17, 18

[12] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel, "mt5: A massively multilingual pre-trained text-to-text transformer," *CoRR*, vol. abs/2010.11934, 2020. [Online]. Available: https://arxiv.org/abs/2010.11934 2

[13] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov.

2019, pp. 6382–6388. [Online]. Available: https://aclanthology.org/D19-1670 2, 20, 37

[14] J. Ács, Á. Kádár, and A. Kornai, "Subword pooling makes a difference," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics, Apr. 2021, pp. 2284–2295. [Online]. Available: https://aclanthology.org/2021.eacl-main.194 2

[15] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 4487–4496. [Online]. Available: https://aclanthology.org/P19-1441 2, 14, 20, 36

[16] F. M. P. del Arco, S. Halat, S. Padó, and R. Klinger, "Multi-task learning with sentiment, emotion, and target detection to recognize hate speech and offensive language," *CoRR*, vol. abs/2109.10255, 2021. [Online]. Available: https://arxiv.org/abs/2109.10255 2

[17] V. Dhananjaya, P. Demotte, S. Ranathunga, and S. Jayasena, "Bertifying sinhala – a comprehensive analysis of pre-trained language models for sinhala text classification," 2022. [Online]. Available: https://arxiv.org/abs/2208.07864 3

[18] M. Artetxe and H. Schwenk, "Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond," *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 597–610, 2019. [Online]. Available: https://aclanthology.org/Q19-1038 4, 11, 35

[19] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. [Online]. Available: https://aclanthology.org/N18-1202 4

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 17351780, nov 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735 4

[21] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR*, vol. abs/1412.3555, 2014. [Online]. Available: http://arxiv.org/abs/1412.3555 4

[22] N. Kalchbrenner, L. Espeholt, K. Simonyan, A. van den Oord, A. Graves, and K. Kavukcuoglu, "Neural machine translation in linear time," *CoRR*, vol. abs/1610.10099, 2016. [Online]. Available: http://arxiv.org/abs/1610.10099 5

[23] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," *CoRR*, vol. abs/1705.03122, 2017. [Online]. Available: http://arxiv.org/abs/1705.03122 5

[24] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," *CoRR*, vol. abs/2005.14165, 2020. [Online]. Available: https://arxiv.org/abs/2005.14165 5

[25] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT sentence embedding," *CoRR*, vol. abs/2007.01852, 2020. [Online]. Available: https://arxiv.org/abs/2007.01852 8

[26] L. Pan, C.-W. Hang, H. Qi, A. Shah, S. Potdar, and M. Yu, "Multilingual BERT post-pretraining alignment," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, Jun. 2021, pp. 210–219. [Online]. Available: https://aclanthology.org/2021.naacl-main.20 8

[27] Z. Chi, L. Dong, F. Wei, N. Yang, S. Singhal, W. Wang, X. Song, X.-L. Mao, H. Huang, and M. Zhou, "InfoXLM: An information-theoretic framework for cross-lingual language model pre-training," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, Jun. 2021, pp. 3576–3588. [Online]. Available: https://aclanthology.org/2021.naacl-main.280 8

[28] S. Cao, N. Kitaev, and D. Klein, "Multilingual alignment of contextual word representations," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=r1xCMyBtPS 8

[29] H. Huang, Y. Liang, N. Duan, M. Gong, L. Shou, D. Jiang, and M. Zhou, "Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 2485–2494. [Online]. Available: https://aclanthology.org/D19-1252 8

[30] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, sep 2020. [Online]. Available: https://doi.org/10.10072Fs11431-020-1647-3 8, 12, 14

[31] Z. Dai, Z. Yang, Y. Yang, J. G. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *CoRR*, vol. abs/1901.02860, 2019. [Online]. Available: http://arxiv.org/abs/1901.02860 9

[32] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 1715–1725. [Online]. Available: https://aclanthology.org/P16-1162 9

[33] A. CONNEAU and G. Lample, "Cross-lingual language model pretraining," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/c04c19c2c2474dbf5f7ac4372c5b9af1-Paper.pdf 9, 10

[34] T. Kudo, "Subword regularization: Improving neural network translation models with multiple subword candidates," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 66–75. [Online]. Available: https://aclanthology.org/P18-1007 9

[35] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018, pp. 66–71. [Online]. Available: https://aclanthology.org/D18-2012 9

[36] S. Wu and M. Dredze, "Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 833–844. [Online]. Available: https://aclanthology.org/D19-1077 12, 20

[37] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune bert for text classification?" in *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 1820, 2019, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2019, p. 194206. [Online]. Available: https://doi.org/10.1007/978-3-030-32381-3_16 12, 13, 20

[38] A. Rosenfeld and J. K. Tsotsos, "Incremental learning through deep adaptation," *CoRR*, vol. abs/1705.04228, 2017. [Online]. Available: http://arxiv.org/abs/1705.04228 12

[39] Z. Wang, W. Bi, Y. Wang, and X. Liu, "Better fine-tuning via instance weighting for text classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7241–7248, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/4709 12

[40] I. Li, P. Sen, H. Zhu, Y. Li, and D. Radev, "Improving cross-lingual text classification with zero-shot instance-weighting," in *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 1–7. [Online]. Available: https://aclanthology.org/2021.repl4nlp-1.1 12

[41] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for NLP," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 2790–2799. [Online]. Available: https://proceedings.mlr.press/v97/houlsby19a.html 13, 37

[42] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," ser. Psychology of Learning and Motivation, G. H. Bower, Ed. Academic Press, 1989, vol. 24, pp. 109–165. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079742108605368 13

[43] B. Ermis, G. Zappella, M. Wistuba, and C. Archambeau, "Memory efficient

continual learning for neural text classification," 2022. [Online]. Available: https://arxiv.org/abs/2203.04640 13

[44] Z. Ke, H. Xu, and B. Liu, "Adapting BERT for continual learning of a sequence of aspect sentiment classification tasks," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, Jun. 2021, pp. 4746–4755. [Online]. Available: https://aclanthology.org/2021.naacl-main.378 13

[45] Z. Ke, B. Liu, H. Wang, and L. Shu, "Continual learning with knowledge transfer for sentiment classification," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 1418, 2020, Proceedings, Part III*. Berlin, Heidelberg: Springer-Verlag, 2020, p. 683698. [Online]. Available: https://doi.org/10.1007/978-3-030-67664-3_41 13

[46] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 328–339. [Online]. Available: https://aclanthology.org/P18-1031 13

[47] D. Hershcovich, S. Frank, H. Lent, M. de Lhoneux, M. Abdou, S. Brandl, E. Bugliarello, L. C. Piqueras, I. Chalkidis, R. Cui, C. Fierro, K. Margatina, P. Rust, and A. Søgaard, "Challenges and strategies in cross-cultural nlp," 2022. [Online]. Available: https://arxiv.org/abs/2203.10020 13

[48] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 440–447. [Online]. Available: https://aclanthology.org/P07-1056 13

[49] M.-W. Chang, L. Ratinov, D. Roth, and V. Srikumar, "Importance of semantic representation: Dataless classification," in *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, ser. AAAI'08. AAAI Press, 2008, p. 830835. 14

[50] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *CoRR*, vol. abs/1802.07569, 2018. [Online]. Available: http://arxiv.org/abs/1802.07569 14

[51] Z. Ke, H. Xu, and B. Liu, "Adapting BERT for continual learning of a sequence of aspect sentiment classification tasks," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, Jun. 2021, pp. 4746–4755. [Online]. Available: https://aclanthology.org/2021.naacl-main.378 14

[52] Z. Ke, B. Liu, H. Wang, and L. Shu, "Continual learning with knowledge transfer for sentiment classification," *ArXiv*, vol. abs/2112.10021, 2020. 14

[53] H. Le, L. Vial, J. Frej, V. Segonne, M. Coavoux, B. Lecouteux, A. Allauzen, B. Crabbé, L. Besacier, and D. Schwab, "Flaubert: Unsupervised language model pre-training for french," *CoRR*, vol. abs/1912.05372, 2019. [Online]. Available: http://arxiv.org/abs/1912.05372 16, 17, 18, 51

[54] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, É. de la Clergerie, D. Seddah, and B. Sagot, "CamemBERT: a tasty French language model," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 7203–7219. [Online]. Available: https://aclanthology.org/2020.acl-main.645 16, 17, 18

[55] M. Polignano, P. Basile, M. de Gemmis, G. Semeraro, and V. Basile, "Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets," in *Proceedings of the Sixth Italian Conference on Computational Linguistics, Bari, Italy, November 13-15, 2019*, ser. CEUR Workshop Proceedings, R. Bernardi, R. Navigli, and G. Semeraro, Eds., vol. 2481. CEUR-WS.org, 2019. [Online]. Available: http://ceur-ws.org/Vol-2481/paper57.pdf 16

[56] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for chinese bert," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021. 16

[57] W. Antoun, F. Baly, and H. M. Hajj, "Arabert: Transformer-based model for arabic language understanding," *CoRR*, vol. abs/2003.00104, 2020. [Online]. Available: https://arxiv.org/abs/2003.00104 16

[58] A. Virtanen, J. Kanerva, R. Ilo, J. Luoma, J. Luotolahti, T. Salakoski, F. Ginter, and S. Pyysalo, "Multilingual is not enough: BERT for finnish," *CoRR*, vol. abs/1912.07076, 2019. [Online]. Available: http://arxiv.org/abs/1912.07076 16, 17

[59] D. Q. Nguyen and A. Tuan Nguyen, "PhoBERT: Pre-trained language models for Vietnamese," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1037–1042. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.92 17, 51

[60] S. Ralethe, "Adaptation of deep bidirectional transformers for Afrikaans language," in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 2475–2478. [Online]. Available: https://aclanthology.org/2020.lrec-1.301 16

[61] S. Wu and M. Dredze, "Are all languages created equal in multilingual BERT?" in *Proceedings of the 5th Workshop on Representation Learning for NLP*. Online: Association for Computational Linguistics, Jul. 2020, pp. 120–130. [Online]. Available: https://aclanthology.org/2020.repl4nlp-1.16 17, 18, 19

[62] D. Kakwani, A. Kunchukuttan, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra, and P. Kumar, "IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 4948–4961. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.445 18, 21, 51

[63] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," *CoRR*, vol. abs/1909.11942, 2019. [Online]. Available: http://arxiv.org/abs/1909.11942 18

[64] U. Khalid, M. O. Beg, and M. U. Arshad, "RUBERT: A bilingual roman urdu BERT using cross lingual transfer learning," *CoRR*, vol. abs/2102.11278, 2021. [Online]. Available: https://arxiv.org/abs/2102.11278 18

[65] J. Ács, D. Lévai, and A. Kornai, "Evaluating transferability of bert models on uralic languages," *arXiv preprint arXiv:2109.06327*, 2021. 19

[66] Y. Kuratov and M. Arkhipov, "Adaptation of deep bidirectional multilingual transformers for russian language," *arXiv preprint arXiv:1905.07213*, 2019. 19

[67] S. Doddapaneni, G. Ramesh, A. Kunchukuttan, P. Kumar, and M. M. Khapra, "A primer on pretrained multilingual language models," *CoRR*, vol. abs/2107.00676, 2021. [Online]. Available: https://arxiv.org/abs/2107.00676 19

[68] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, and M. Johnson, "XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization," *CoRR*, vol. abs/2003.11080, 2020. [Online]. Available: https://arxiv.org/abs/2003.11080 20

[69] A. Lauscher, V. Ravishankar, I. Vulić, and G. Glavaš, "From zero to hero: On the limitations of zero-shot language transfer with multilingual Transformers," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 4483–4499. [Online]. Available: https://aclanthology.org/2020.emnlp-main.363 19

[70] G. Aguilar, S. Kar, and T. Solorio, "LinCE: A centralized benchmark for linguistic code-switching evaluation," in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 1803–1813. [Online]. Available: https://aclanthology.org/2020.lrec-1.223 19

[71] S. Groenwold, S. Honnavalli, L. Ou, A. Parekh, S. Levy, D. Mirza, and W. Y. Wang, "Evaluating the role of language typology in transformer-based multilingual text classification," *CoRR*, vol. abs/2004.13939, 2020. [Online]. Available: https://arxiv.org/abs/2004.13939 19

[72] A. Ebrahimi, M. Mager, A. Oncevay, V. Chaudhary, L. Chiruzzo, A. Fan, J. Ortega, R. Ramos, A. Rios, I. Vladimir, G. A. Giménez-Lugo, E. Mager, G. Neubig, A. Palmer, R. A. C. Solano, N. T. Vu, and K. Kann, "Americasnli: Evaluating zero-shot natural language understanding of pretrained multilingual models in truly low-resource languages," *CoRR*, vol. abs/2104.08726, 2021. [Online]. Available: https://arxiv.org/abs/2104.08726 20

[73] R. Litschko, I. Vulic, S. P. Ponzetto, and G. Glavas, "Evaluating multilingual text encoders for unsupervised cross-lingual retrieval," *CoRR*, vol. abs/2101.08370, 2021. [Online]. Available: https://arxiv.org/abs/2101.08370 20

[74] A. Ebrahimi and K. Kann, "How to adapt your pretrained multilingual model to 1600 languages," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 4555–4567. [Online]. Available: https://aclanthology.org/2021.acl-long.351 20

[75] Y. Fujinuma, J. Boyd-Graber, and K. Kann, "Match the script, adapt if multilingual: Analyzing the effect of multilingual pretraining on cross-lingual transferability," 2022. [Online]. Available: https://arxiv.org/abs/2203.10753 20

[76] J. Ma and L. Li, "Data augmentation for chinese text classification using back-translation," *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012039, nov 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1651/1/012039 20

[77] M. Bayer, M. Kaufhold, and C. Reuter, "A survey on data augmentation for text classification," *CoRR*, vol. abs/2107.03158, 2021. [Online]. Available: https://arxiv.org/abs/2107.03158 20

[78] Y. Meng, Y. Zhang, J. Huang, C. Xiong, H. Ji, C. Zhang, and J. Han, "Text classification using label names only: A language model self-training approach," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 9006–9017. [Online]. Available: https://aclanthology.org/2020.emnlp-main.724 21

[79] S. H. Muhammad, D. I. Adelani, S. Ruder, I. S. Ahmad, I. Abdulmumin, B. S. Bello, M. Choudhury, C. C. Emezue, S. A. Salahudeen, A. Anuoluwapo, A. Jeorge, and P. Brazdil, "Naijasenti: A nigerian twitter sentiment corpus for multilingual sentiment analysis," *CoRR*, vol. abs/2201.08277, 2022. [Online]. Available: https://arxiv.org/abs/2201.08277 21

[80] D. Aggarwal, V. Gupta, and A. Kunchukuttan, "Indicxnli: Evaluating multilingual inference for indian languages," 2022. [Online]. Available: https://arxiv.org/abs/2204.08776 21

[81] B. Müller, Y. Elazar, B. Sagot, and D. Seddah, "First align, then predict: Understanding the cross-lingual ability of multilingual BERT," *CoRR*, vol. abs/2101.11109, 2021. [Online]. Available: https://arxiv.org/abs/2101.11109 21

[82] J. Singh, B. McCann, R. Socher, and C. Xiong, "BERT is not an interlingua and the bias of tokenization," in *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 47–55. [Online]. Available: https://aclanthology.org/D19-6106 22

[83] S. Wu and M. Dredze, "Do explicit alignments robustly improve multilingual encoders?" in *Proceedings of the 2020 Conference on Empirical Methods*

in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 4471–4482. [Online]. Available: https://aclanthology.org/2020.emnlp-main.362 22

[84] Y. Wang, W. Che, J. Guo, Y. Liu, and T. Liu, "Cross-lingual BERT transformation for zero-shot dependency parsing," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5721–5727. [Online]. Available: https://aclanthology.org/D19-1575 22

[85] X. Chen, H. Fan, R. B. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *CoRR*, vol. abs/2003.04297, 2020. [Online]. Available: https://arxiv.org/abs/2003.04297 22, 51

[86] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proceedings of Machine Translation Summit X: Papers*, Phuket, Thailand, Sep. 13-15 2005, pp. 79–86. [Online]. Available: https://aclanthology.org/2005.mtsummit-papers.11 22

[87] P. Efimov, L. Boytsov, E. Arslanova, and P. Braslavski, "The impact of cross-lingual adjustment of contextual word representations on zero-shot transfer," 2022. [Online]. Available: https://arxiv.org/abs/2204.06457 22

[88] Z. Liu, G. I. Winata, A. Madotto, and P. Fung, "Preserving cross-linguality of pre-trained models via continual learning," in *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 64–71. [Online]. Available: https://aclanthology.org/2021.repl4nlp-1.8 23

[89] D. Lopez-Paz and M. Ranzato, "Gradient episodic memory for continual learning," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 64706479. 23

[90] F. Å. Nielsen, "A new ANEW: evaluation of a word list for sentiment analysis in microblogs," *CoRR*, vol. abs/1103.2903, 2011. [Online]. Available: http://arxiv.org/abs/1103.2903 23

[91] S. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

Papers). Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 174–184. [Online]. Available: https://aclanthology.org/P18-1017 23, 39, 43

[92] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550 43

[93] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA), May 2010. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf 23

[94] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 06 2011. [Online]. Available: https://doi.org/10.1162/COLI_a_00049 23

[95] C. Musto, G. Semeraro, and M. Polignano, "A comparison of lexicon-based approaches for sentiment analysis of microblog posts." in *DART@ AI* IA*. Citeseer, 2014, pp. 59–68. 23

[96] B. Shin, T. Lee, and J. D. Choi, "Lexicon integrated CNN models with attention for sentiment analysis," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 149–158. [Online]. Available: https://aclanthology.org/W17-5220 23

[97] A. Kumar, D. Kawahara, and S. Kurohashi, "Knowledge-enriched two-layered attention network for sentiment analysis," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 253–258. [Online]. Available: https://aclanthology.org/N18-2041 23

[98] K. Margatina, C. Baziotis, and A. Potamianos, "Attention-based conditioning methods for external knowledge integration," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy:

Association for Computational Linguistics, Jul. 2019, pp. 3944–3951. [Online]. Available: https://aclanthology.org/P19-1385

[99] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/12048 23

[100] V. Suresh and D. C. Ong, "Using knowledge-embedded attention to augment pre-trained language models for fine-grained emotion recognition," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, pp. 1–8. 23, 39, 51

[101] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "SentiLARE: Sentiment-aware language representation learning with linguistic knowledge," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 6975–6988. [Online]. Available: https://aclanthology.org/2020.emnlp-main.567 23, 40

[102] P. Zhong, D. Wang, and C. Miao, "Knowledge-enriched transformer for emotion detection in textual conversations," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 165–176. [Online]. Available: https://aclanthology.org/D19-1016 23

[103] N. de Silva, "Survey on publicly available sinhala natural language processing tools and research," *CoRR*, vol. abs/1906.02358, 2019. [Online]. Available: http://arxiv.org/abs/1906.02358 23

[104] S. Fernando, S. Ranathunga, S. Jayasena, and G. Dias, "Comprehensive part-of-speech tag set and SVM based POS tagger for Sinhala," in *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)*. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 173–182. [Online]. Available: https://aclanthology.org/W16-3718 23

[105] S. Manamini, A. Ahamed, R. Rajapakshe, G. Reemal, S. Jayasena, G. Dias, and S. Ranathunga, "Ananya - a named-entity-recognition (ner) system for sinhala language," in *2016 Moratuwa Engineering Research Conference (MERCon)*, 2016, pp. 30–35. 23

[106] K. Kumarasinghe, G. Dias, and I. Herath, "Sinmorphy: A morphological analyzer for the sinhala language," in *2021 Moratuwa Engineering Research Conference (MERCon)*, 2021, pp. 681–686. 23

[107] D. Upeksha, C. Wijayarathna, M. Siriwardena, L. Lasandun, C. Wimalasuriya, N. de Silva, and G. Dias, "Implementing a corpus for sinhala language," 01 2015. 23

[108] P. Nanayakkara and S. Ranathunga, "Clustering sinhala news articles using corpus-based similarity measures," *2018 Moratuwa Engineering Research Conference (MERCon)*, pp. 437–442, 2018. 24

[109] R. Weerasinghe, D. Herath, and V. Welgama, "Corpus-based Sinhala lexicon," in *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*. Suntec, Singapore: Association for Computational Linguistics, Aug. 2009, pp. 17–23. [Online]. Available: https://aclanthology.org/W09-3403 24

[110] S. Thillainathan, S. Ranathunga, and S. Jayasena, "Fine-tuning self-supervised multilingual sequence-to-sequence models for extremely low-resource nmt," in *2021 Moratuwa Engineering Research Conference (MERCon)*. IEEE, 2021, pp. 432–437. 24

[111] S. Gallege, "Analysis of sinhala using natural language processing techniques," *University of Wisconsin*, 2010. 24

[112] K. Lakmali and P. S. Haddela, "Effectiveness of rule-based classifiers in sinhala text categorization," in *2017 National Information Technology Conference (NITC)*. IEEE, 2017, pp. 153–158. 24

[113] S. Gunasekara and P. S. Haddela, "Context aware stopwords for sinhala text classification," in *2018 National Information Technology Conference (NITC)*. IEEE, 2018, pp. 1–6. 24

[114] N. P. K. Medagoda, "Framework for sentiment classification for morphologically rich languages: A case study for sinhala," Ph.D. dissertation, Auckland University of Technology, 2017. 24

[115] P. Chathuranga, S. Lorensuhewa, and M. Kalyani, "Sinhala sentiment analysis using corpus based sentiment lexicon," in *2019 19th international conference on advances in ICT for emerging regions (ICTer)*, vol. 250. IEEE, 2019, pp. 1–7. 24

[116] P. Jayasuriya, B. Kumarasinghe, S. Ekanayake, R. Munasinghe, S. Thelijjagoda, and I. Weerasinghe, "Sentiment classification of sinhala content in social media," 09 2020. 24

[117] S. Ranathunga and I. U. Liyanage, "Sentiment analysis of sinhala news comments," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 4, pp. 1–23, 2021. 24

[118] P. Demotte, L. Senevirathne, B. Karunanayake, U. Munasinghe, and S. Ranathunga, "Sentiment analysis of sinhala news comments using sentence-state lstm networks," in *2020 Moratuwa Engineering Research Conference (MERCon)*. IEEE, 2020, pp. 283–288. 24

[119] Y. Zhang, Q. Liu, and L. Song, "Sentence-state lstm for text representation," *arXiv preprint arXiv:1805.02474*, 2018. 24

[120] L. Senevirathne, P. Demotte, B. Karunanayake, U. Munasinghe, and S. Ranathunga, "Sentiment analysis for sinhala language using deep learning techniques," *arXiv preprint arXiv:2011.07280*, 2020. 24, 27, 31, 35, 42

[121] H. Rathnayake, J. Sumanapala, R. Rukshani, and S. Ranathunga, "Adapter-based fine-tuning of pre-trained multilingual language models for code-mixed and code-switched text classification," *Knowledge and Information Systems*, vol. 64, no. 7, pp. 1937–1966, Jul 2022. [Online]. Available: https://doi.org/10.1007/s10115-022-01698-1 24

[122] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017. 25, 44

[123] N. de Silva, "Sinhala text classification: Observations from the perspective of a resource poor language," 06 2015. 27

[124] D. Sachintha, L. Piyarathna, C. Rajitha, and S. Ranathunga, "Exploiting parallel corpora to improve multilingual embedding based document and sentence alignment," *CoRR*, vol. abs/2106.06766, 2021. [Online]. Available: https://arxiv.org/abs/2106.06766 27

[125] D. Upeksha, C. Wijayarathna, M. Siriwardena, L. Lasandun, C. Wimalasuriya, N. de Silva, and G. Dias, "Implementing a corpus for sinhala language," 01 2015. 28

[126] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019. 30

[127] A. Sarkar, S. Reddy, and R. S. Iyengar, "Zero-shot multilingual sentiment analysis using hierarchical attentive network and bert," in *Proceedings of the 2019*

3rd International Conference on Natural Language Processing and Information Retrieval, 2019, pp. 49–56. 36

[128] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, and E. Hovy, "A survey of data augmentation approaches for nlp," *arXiv preprint arXiv:2105.03075*, 2021. 37

[129] M. Wortsman, G. Ilharco, M. Li, J. W. Kim, H. Hajishirzi, A. Farhadi, H. Namkoong, and L. Schmidt, "Robust fine-tuning of zero-shot models," *CoRR*, vol. abs/2109.01903, 2021. [Online]. Available: https://arxiv.org/abs/2109.01903 37

[130] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 380–385. [Online]. Available: https://aclanthology.org/N19-1035 41

[131] E. Wallace, S. Feng, N. Kandpal, M. Gardner, and S. Singh, "Universal adversarial triggers for attacking and analyzing NLP," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 2153–2162. [Online]. Available: https://aclanthology.org/D19-1221 42

[132] C. Potts, Z. Wu, A. Geiger, and D. Kiela, "Dynasent: A dynamic benchmark for sentiment analysis," *CoRR*, vol. abs/2012.15349, 2020. [Online]. Available: https://arxiv.org/abs/2012.15349 42

[133] A. Hande, S. U. Hegde, R. Priyadharshini, R. Ponnusamy, P. K. Kumaresan, S. Thavareesan, and B. R. Chakravarthi, "Benchmarking multi-task learning for sentiment analysis and offensive language identification in under-resourced dravidian languages," *CoRR*, vol. abs/2108.03867, 2021. [Online]. Available: https://arxiv.org/abs/2108.03867 43

[134] K. I. Islam, M. S. Islam, and M. R. Amin, "Sentiment analysis in bengali via transfer learning using multi-lingual BERT," *CoRR*, vol. abs/2012.07538, 2020. [Online]. Available: https://arxiv.org/abs/2012.07538 43

[135] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html 47

[136] P. Littell, D. R. Mortensen, K. Lin, K. Kairis, C. Turner, and L. Levin, "URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 8–14. [Online]. Available: https://aclanthology.org/E17-2002 49