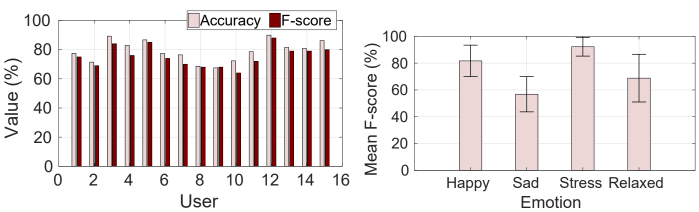
imbalance by applying Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is designed to re-sample the class (emotion state) with the least number of instances so that all classes have almost equal samples. After balancing the dataset, we have in total 8301 typing sessions (on average 553 sessions per user, std. dev. 504.2). We show the distribution of these typing sessions tagged with different emotion labels before and after applying SMOTE in Figure 2.

4 EVALUATION

4.1 Model Performance

We perform 10-fold cross validation to evaluate the model. We report the user-wise classification accuracy and F-score in Figure 3a. We obtain an average accuracy of 80% (std dev. 7.1%). The minimum accuracy obtained across all users is 69%, while the highest one is 90%. We obtain an average F-score of 75% (std dev. 7%). We also report the state-wise F-score in Figure 3b. It is observed that all emotions except *sad* are having a F-score more than 70%, while *stressed* is detected with an F-score of 92%.



(a) User-wise Accuracy, F-score **(b) Emotion-wise F-score**

Figure 3: Emotion classification performance of the proposed DNN model. Error bar indicates standard deviation.

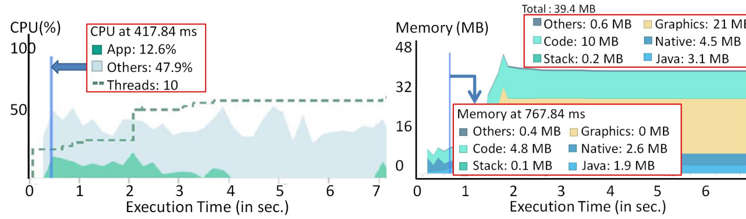
4.2 Resource Overhead

In order to measure resource overhead, we deploy the models on smartphone. We use Tensorflow to generate the deployable model for Android system. We deploy the models on a Moto G2 phone. We show the CPU utilization and memory consumption of the app in Figure 4a, 4b respectively. It is observed that peak CPU utilization is less than 15%, whereas the cumulative memory consumption is less than 40 MB.

Parameter	Mean	Std dev
Model size (in KB)	3.7	0.0
Load time (in msec.)	240.5	12.6
Inference time (in msec.)	3.2	9.6

Table 2: User-wise mean and std deviation of trained model size, model load time and inference time

We also measure trained model size and time required to load the model file for every user. We compute the average model size and average load time for every user and report the same in Table 2.



(a) CPU utilization **(b) Memory consumption**

Figure 4: Resource overhead in terms of CPU utilization and memory consumption (peak CPU utilization less than 15% and cumulative memory consumption less than 40 MB)

Similarly, we compute the average inference time for a test instance. We randomly select 20% samples from every user and use corresponding model to predict the outcome. The time required to infer the emotion for a single test instance is considered as inference time. We note the average inference time for all test instances from every user in Table 2 also. We observe a high standard deviation in inference time. This is primarily because that first inference usually takes more time, later on once the model file is loaded, the inference time is reduced by large amount, thus resulting in high standard deviation.

5 CONCLUSION

In this paper, we investigate the feasibility of executing DNN models on smartphone to determine multiple emotion states based on typing. We develop a MLP based personalized DNN model based on typing features to determine four emotion states (*happy, sad, stressed, relaxed*) and deploy the same on smartphone. The evaluation of the model reveals that it is possible to determine these states with an average accuracy of 80% without major resource overhead.

REFERENCES

[1] Bokai Cao, Lei Zheng, Chenwei Zhang, Philip S Yu, Andrea Piscitello, John Zulueta, Olu Ajilore, Kelly Ryan, and Alex D Leow. 2017. Deepmood: modeling mobile phone typing dynamics for mood detection. In *ACM SIGKDD*.

[2] Surjya Ghosh, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2017a. Evaluating effectiveness of smartphone typing as an indicator of user emotion. In *IEEE ACII*.

[3] Surjya Ghosh, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2017b. TapSense: Combining Self-Report Patterns and Typing Characteristics for Smartphone based Emotion Detection. In *ACM MobileHCI*.

[4] Natasha Jaques, Sara Taylor, Akane Sano, Rosalind Picard, and others. 2017. Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing*. 17–33.

[5] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *ACM Mobisys*.

[6] James A Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178.