

# On the Influence of Beta Cell Granule Counting for Classification in Type 1 Diabetes

Lea Bogensperger<sup>1</sup>, Marc Masana<sup>1,2</sup>, Filip Ilic<sup>1</sup>, Dagmar Kolb<sup>3,4</sup>, and Thomas Pock<sup>1</sup>

**Abstract**—Patients suffering with type 1 diabetes show a major reduction of  $\beta$ -cells within their pancreas. By analyzing tissue samples containing granules – the insulin-producing units within the  $\beta$ -cells – we aim to gather more information on the respective healthy and diabetic phenotypes, which could lead to further understanding the pathogenesis of the disease. To this end, we use a deep learning approach to investigate whether assumptions on the pathological status can be made based on electron micrograph images of  $\beta$ -cells. To support the decision-making process we explore whether estimating the number of granules can be used to aid in discriminating healthy from diabetic samples. Furthermore, we demonstrate that multi-task and transfer learning strategies can lead to more accurate predictions. Finally, this work intends to contribute to a more in-depth understanding of the structural mechanisms in type 1 diabetes, which is essential to design better approaches to a tailored treatment.

## I. INTRODUCTION

Type 1 diabetes is an autoimmune disease that is characterized by the destruction of insulin-producing  $\beta$ -cells [2]. The body’s immune system attacks its own essential insulin-producing mechanism within the pancreas thus impeding a normal blood glucose regulation after food intake. The main participants in this process are the granules of  $\beta$ -cells that can be found within the islets of Langerhans in the endocrine pancreas. They are composed of an insulin-producing core and surrounded by a less dense halo, as explained in [11]. The author further describes how the granules are suspected to be the focus of the autoimmune-mediated  $\beta$ -cell destruction due to immunogenic targets. An example of such granules is shown in Figure 1, where part of a healthy and a non-obese diabetic (NOD) mouse are shown. Following the attack of the immune system, the overall number of granules contained in the  $\beta$ -cells gets drastically decreased and the remaining  $\beta$ -cells are stressed in trying to maintain insulin supply [3].

By visually inspecting electron micrograph slices such as in Figure 1, it is at first glance not trivial to discriminate the microscopic images between healthy and NOD. Therefore, an important research question deals with finding features that allow for distinguishing between  $\beta$ -cells in healthy and NOD mice. The number of granules that can be found within  $\beta$ -cells seems to be a strong indicator. Furthermore, there are hypotheses in the medical domain regarding structural

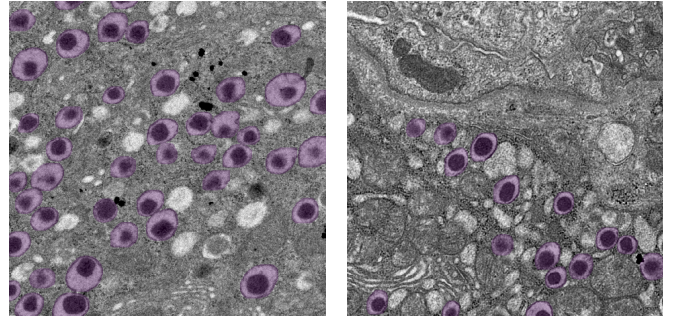


Fig. 1. Slices revealing  $\beta$ -cells of a healthy (left) and a diabetic NOD (right) mouse. Numerous granules, colored in purple, are visible in both samples. They are characterized by their circular structure containing a dark insulin-producing core surrounded by its vesicle characterized by its typical white halo.

changes of the entire  $\beta$ -cell [13] which can result in changed appearance of the granules regarding shape and size.

We therefore propose a deep learning based approach to discriminate between slices containing  $\beta$ -cell granules of healthy and NOD mice. Our focus is to learn a classification system that can distinguish between healthy and early stage type 1 diabetes in electron micrograph images of  $\beta$ -cells with their insulin-producing granules. Further, fueled by the interesting findings that the number of granules seems to be able to distinguish to an extent between healthy and NOD slices (see Section IV), we pursue to train a granule-counting system that for a given  $\beta$ -cell slice learns to count the number of granules present in the sample.

To summarize, our main contributions are:

- we demonstrate the applicability of a deep learning based approach in classifying electron micrograph images as either healthy or NOD,
- we show that a network can also be trained to estimate the number of insulin-producing granules in a given input image, and explore the expressiveness of its features for classification,
- we explore multi-task and transfer learning strategies to combine classification and counting to help in distinguishing healthy from NOD samples.

## II. RELATED WORK

To the best of our knowledge, deep learning approaches have not yet been explored in better understanding disease formation and progression of type 1 diabetes at the level of  $\beta$ -cell granules.

<sup>1</sup> Graz University of Technology, Austria.

<sup>2</sup> Silicon Austria Labs, TU-Graz SAL DES Lab, Austria.

<sup>3</sup> Core Facility Ultrastructure Analysis, Graz, Austria.

<sup>4</sup> Medical University of Graz, Austria.

Corresponding Author: lea.bogensperger@icg.tugraz.at

Interestingly, a similar idea was proposed by Zhang et al. [19], where the authors investigated monkeys with type 2 diabetes with Metabolic syndrome. After using binary segmentation followed by a watershed transform, they were able to obtain granule instances allowing further post-processing on important features regarding the granules appearances such as radius and size of granular core and vesicle.

Regarding type 1 diabetes, extensive research can be found that is focused on predicting development of diabetes by means of machine learning, which was partly also analyzed by a survey conducted recently [16]. Tripathi et al. [14] and Xue et al. [17] aim to predict the development of type 1 diabetes by incorporating risk factors such as blood pressure and blood glucose level by using Random Forests and Support Vector Machines based on statistical measures such as accuracy and precision. On the other hand, Zaitcev et al. [18] train a neural network to model long term average glucose levels from 5-12 weeks of daily measurements whereas Alfian et al. [1] focus on modeling blood glucose levels in the next 30-60 minutes to combat hypoglycemia, a condition with too low blood glucose levels that can become dangerous for affected patients. None of these approaches have been applied to electron micrograph images, which are not applicable in situ in clinical practice.

These feature-based approaches are obviously very relevant in practice to predict the risk of patients developing type 1 diabetes or to model blood glucose levels of patients that already suffer from the disease. Deep learning techniques can be very helpful to enhance prediction and diagnosis of disease development and progression and thus improve living conditions of those affected. Indeed, these approaches can also be used in the process of better understanding the nature of type 1 diabetes and the mechanisms that are triggered at the levels of insulin-producing granules within the  $\beta$ -cells. However, research on this data can hardly be carried out on humans since the tissue cannot be harvested from living humans and donors from deceased patients are often at a stage where the disease has progressed further and no granules are left within the  $\beta$ -cells. Therefore, conducting this research on mice is a well-suited alternative that allows for results to advance the process of better understanding the pathogenesis of type 1 diabetes.

### III. METHODOLOGY

#### A. Classification and Granule Counting

Discrimination between healthy and NOD tomographic  $\beta$ -cell images can be defined as a binary classification problem. We propose to learn an encoder – or feature extractor – which provides a feature representation of the input images onto a latent space that can be used to solve the classification and granule counting tasks (see Figure 2). In the proposed setting, data is scarce due to its acquisition and annotation costs. Therefore, we propose to use ResNet-18 [5] initialized with pre-trained weights on ImageNet [4] as the encoder. Pre-trained models on large datasets have shown to provide good feature representations that allow for a more robust initialization and faster training when fine-tuning on small

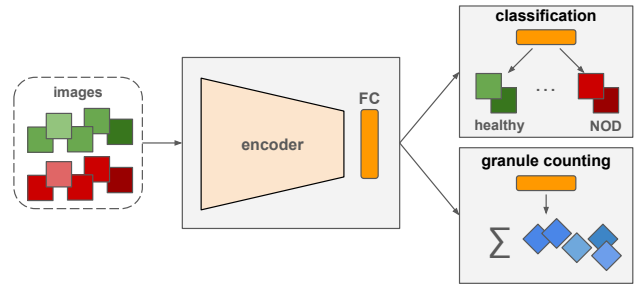


Fig. 2. Proposed architecture to learn classification and/or granule counting.

tasks [8]. Since the output feature space of those pre-trained models are usually of high dimensionality, we propose to include a fully-connected layer between the backbone and the classifiers to enforce a dimensionality reduction. The objective function that is minimized to learn the network parameters is the binary cross-entropy loss

$$\mathcal{L}_C(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^N (y_n \log(\hat{y}_n) + (1 - y_n) \log(1 - \hat{y}_n)), \quad (1)$$

where  $y_n$  is the true label and  $\hat{y}_n$  the predicted label for each sample  $n$ . Meanwhile, for counting the number of granules present in an input image the problem is framed as a regression task. Therefore the backbone of the feature encoder is appended with a fully-connected layer, where the network's output is the estimated number of granules. In this case the mean squared error is used as the objective function during learning

$$\mathcal{L}_R(y, \hat{y}) = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2, \quad (2)$$

where a strong penalization for higher deviations in the predicted number of granules from the groundtruth is desirable.

#### B. Multi-task and Transfer Learning

We assume that granule counting can help discriminate between healthy and NOD samples – we show that this assumption holds in Section IV-B. Traditionally, there are two popular strategies in which both tasks can be incorporated into the learning process to benefit each other during training. The first is multi-task learning [12], where both tasks are learned together at the same time on different heads from the shared fully-connected layer. To this end, the loss function is the weighted combination of the binary cross-entropy loss and the mean-squared error

$$\mathcal{L}_{\text{joint}}(y, \hat{y}) = \lambda \mathcal{L}_C(y, \hat{y}) + (1 - \lambda) \mathcal{L}_R(y, \hat{y}). \quad (3)$$

where  $\lambda \in (0, 1)$  is the trade-off between the two tasks. Learning both tasks ensures that the learned, shared feature space contains meaningful representations for both tasks, either helping each other with shared discriminative features, or with regularization over the capacity and importance of the learned features.

Further, as mentioned above, fine-tuning over pre-trained models has been shown to be a useful means of transferring

good mid-level representations [8]. This idea can be extended beyond models trained on large datasets to smaller tasks also via transfer learning [9]. Therefore, we propose that the backbone and fully-connected layer are trained on the regression task to then use the resulting model as an initialization for learning the classification task – each task learned on the corresponding head. The underlying intuition is that features which have been learned during granule counting can be beneficial for the classification process since these tasks are shown to be related by the discriminative capacity of the number of granules for healthy and NOD samples.

### C. Evaluation Metrics

The accuracy for binary classification is given by

$$\mathcal{A}(y, \hat{y}) = \frac{1}{N} \sum_{n=1}^N \delta(y_n, \hat{y}_n), \quad (4)$$

with  $\delta(i, j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else,} \end{cases}$

where  $\delta$  is the indicator function. To quantitatively evaluate the potential of the estimated number of granules to classify between healthy and NOD samples, another evaluation criterion is required. We propose a new metric to compute regression accuracy which searches for the optimal threshold  $\theta^*$  of granules per sample. This threshold separates below and above which samples are predicted to be NOD or healthy, respectively. On training data, it is found by

$$\theta^* = \operatorname{argmax}_{\theta} \mathcal{A}(y, \hat{y}_{\theta}), \quad (5)$$

where  $\hat{y}_{\theta}$  is the predicted label based on the threshold  $\theta$ . Then, once fixed, samples from the test set are classified solely based on the number of counted granules, whereupon the accuracy in Equation 4 can be computed.

### D. Dataset

Our underlying objective to the presented methodology is to gain novel insights into the mechanisms of  $\beta$ -cell loss via use of neural networks capable of analyzing images containing insulin-producing  $\beta$ -cells. Autophagy, apoptosis, and endoplasmic reticulum stress are indicators of stress responses in NOD  $\beta$ -cells after immune system attacks, which are expected to have an influence on the produced insulin through a changed appearance of the granules themselves or through the overall number of granules that are available for insulin production. Therefore, we created a dataset to obtain electron micrographs of pancreata of healthy and NOD mice at different ages. Ultrathin sections (70 nm) are stained with platine blue and lead citrate. Non-overlapping areas of pancreatic islets are then pre-selected and analysed by transmission electron microscopy (FEI Tecnai G2) at 120kV. All images were acquired with a magnification of  $2,500\times$  and have a resolution of  $1024 \times 1024$ .

The dataset consists of 362 tomographic images from  $\beta$ -cells of healthy C57BL/6J and NOD mice (3/3), which are widely used in type 1 diabetic research [7], [10]. An exemplary image of each group is shown in Figure 1. The

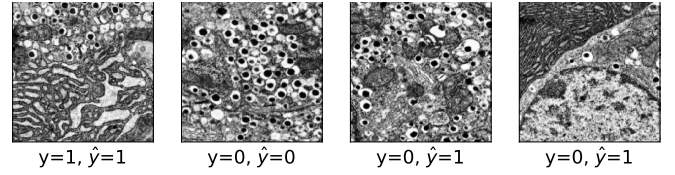


Fig. 3. Examples of correctly classified (left) and misclassified (right) samples from the test set, where  $y$  denotes the true class label, and  $\hat{y}$  the predicted label. Class labels 0 and 1 denote healthy and NOD, respectively.

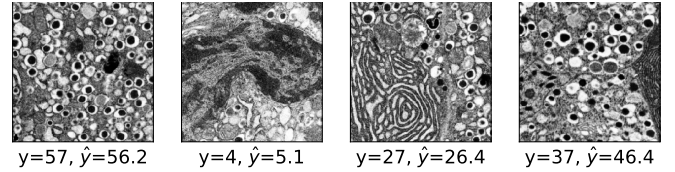


Fig. 4. Examples of granule counting estimation on samples from the test set, where  $y$  denotes the groundtruth number of granules, and  $\hat{y}$  the predicted number. The rightmost image is the one with the highest counting error on test, probably due to the presence of immature granules and granules from  $\alpha$ -cells, which were not included in the training data.

slices are non-overlapping and cover several distinct  $\beta$ -cells within each pancreas. The dataset is split into 200 training images and 62 testing images, while ensuring that available data from each individual mouse is either part of the train or the test set.

### E. Experimental Setup

The chosen encoder backbone is ResNet-18 [5] pre-trained on ImageNet [4]. On top of the encoder, a shared fully-connected layer of size  $512 \times 256$  is added to reduce the dimensionality of the latent space before the regression and the classification head. By definition of the losses in Section III-A, the first head has 2 outputs representing the binary classification probabilities of healthy and NOD, while the second head has a single output which predicts the number of granules present in the image.

Input images are pre-processed in the same fashion as the encoder backbone was pre-trained. Images are resized to  $224 \times 224$  followed by standard data augmentation techniques consisting of horizontal and vertical flips. In addition, histogram equalization is applied to all images to compensate for uneven illumination and differences in microscopic settings. During training, we use Adam optimization [6] with an initial learning rate of  $1e-4$  for classification and  $1e-3$  for counting regression, and exponential decay rates of  $\beta_1 = 0.9, \beta_2 = 0.999$ .

## IV. RESULTS

### A. Classification

Classification results on the test set achieved an accuracy of 93.54%. Figure 3 shows examples of correctly and misclassified samples, where class 0 and class 1 indicate healthy and NOD, respectively.

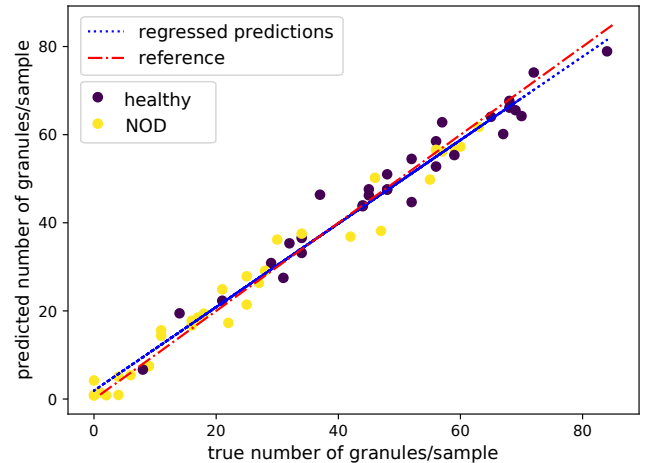
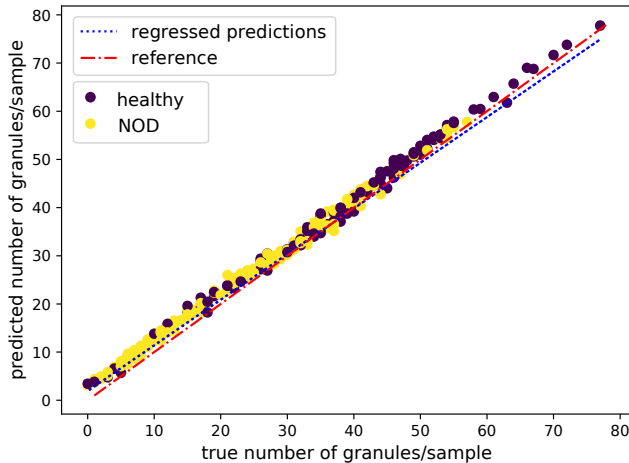


Fig. 5. For both the train (left) and test set (right) in the regression task, the predicted vs. the groundtruth number of granules are displayed with their true class labels. The tendency that healthy samples contain more granules than NOD can be visually observed in both cases.

### B. Granule Counting

Training the model on granule counting yields a mean absolute error of 2.78. Examples of granule estimation on the test set are shown in Figure 4. Deviations from the groundtruth number of granules might arise due to the fact that  $\beta$ -cells contain other cellular components with circular structures that may resemble granules and are thus hard to distinguish. The maximum absolute error obtained across all test images is 9.38, shown at the rightmost of Figure 4. This image contains some immature granules and  $\alpha$ -cell granules, which the network has not learned to discriminate from  $\beta$ -cell granules due to the limited presence of  $\alpha$ -cell granules during training. Furthermore, it has to be noted that it is challenging even for human observers to agree on the exact number of granules, since some imaged granules are out of focus or subject to some microscopic acquisition artefacts. Therefore, we argue that the predicted number of granules is within reason to assume that the model has learned the task properly.

Figure 5 shows the predicted number of granules per sample with respect to the groundtruth. We observe a similar light underestimation on both train and test for images with more granules and a neglectable overestimation for images with lower granule count. Furthermore, each sample is labeled with its true class label, where for both train and test data a relation between a larger/smaller number of granules and a healthy class label 0/NOD class label 1 can be observed.

### C. Granule Counting for Classification

As seen in the previous experiment, using the number of estimated granules in a given sample can be an approximate indicator to classify it as either healthy or NOD. Therefore, we apply the optimal threshold search proposed in Section III-C to the granule counting model results on training data. We provide the curve profile of the training data and the corresponding optimal threshold  $\theta^*$  in Figure 6.

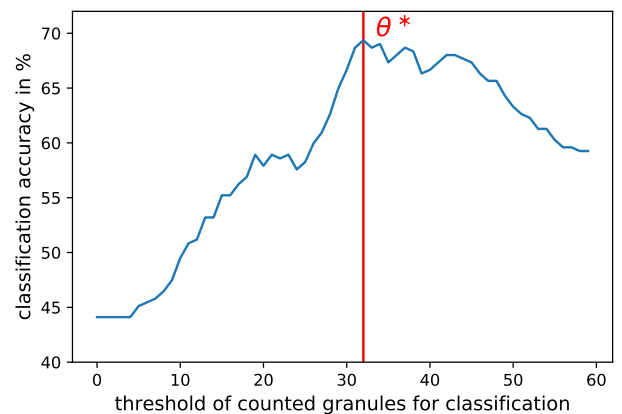


Fig. 6. Evaluation of the classification accuracy on samples, where solely the regression task of counting granules has been trained. Classification accuracy is obtained by thresholding at  $\theta$  number of granules per slice. Above that threshold value a sample is considered as healthy. Setting the threshold to  $\theta^* = 32$  granules/slice yields the maximum classification accuracy of 69.36% within the training data.

Classification results on the training data yield a maximum accuracy of 69.36% ( $\theta^* = 32$ ). With the learned  $\theta^* = 32$  applied to the test data, an accuracy of 74.19% is achieved – that is separating solely by granule count. The fact that the accuracy on test is higher than train highlights how the model was not directly trained for the classification task, although it can perform significantly well on it. Doing a Welch’s t-test [15] on the estimated number of granules per sample based on the true class label yields a t-statistic of 4.96 with a corresponding p-value of  $6.3 \times 10^{-6}$ , which further emphasizes the strong significant difference in number of granules between healthy and NOD.

### D. Joint Granule Counting and Classification

Following the results presented above, granule counting can help distinguish between healthy and NOD images. As

TABLE I  
SUMMARY OF THE RESULTS ON THE INFLUENCE OF  $\beta$ -CELL GRANULE COUNTING FOR CLASSIFICATION.

|   | test accuracy |
|---|---------------|
| only classification                                 | 93.54%        |
| only counting regression                            | 74.19%        |
| joint ( $\lambda = 0.8$ )                           | 95.16%        |
| transfer learning (counting $\rightarrow$ classif.) | 96.77%        |

mentioned in Section III-B, we can then exploit this synergy to propose a joint training of both tasks. We simultaneously train both tasks under the same architecture using Equation 3 and with the balancing parameter  $\lambda$ . A grid search over  $\lambda \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$  shows that the best result on the train set is obtained at  $\lambda = 0.8$ . Setting  $\lambda = 0$  or  $\lambda = 1$  yields the baselines of training only regression or classification, respectively. As expected, classification is given more emphasis since it is the main task at which we evaluate, and we hypothesize that counting granules helps in classification in the sense of a regularizer that learns a better feature representation. This joint multi-task training achieves a classification accuracy of 95.16%, which improves with respect to solely classification or regression, as summarized in Table I.

#### E. Transfer Learning

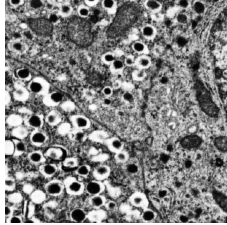
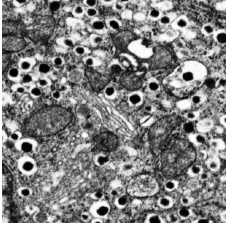
Finally, we apply transfer learning from the network trained on granule counting by training classification on top of it. This strategy achieves an accuracy of 96.77%, which surpasses all other results (see Table I). Although the performance of transfer learning is slightly above that of the joint training, we can not exclude that an unexplored  $\lambda$  trade-off exists, which provides a better accuracy. In conclusion, both strategies which exploit the knowledge provided by counting granules when learning to discriminate between healthy and NOD samples outperform the strategies which only rely on a single task.

Table II shows an instance where granule counting has a positive impact on the healthy/NOD classification for both joint training and transfer learning. Both healthy samples were predicted as NOD by using only classification. Due to granule counting, which for both samples is rather high, the prediction changes to the correct class when evaluated on the joint or transfer learning models. In the case of joint training, we also provide the prediction for counting.

#### V. CONCLUSION AND FUTURE WORK

We use a deep learning based approach to achieve high classification accuracy when discriminating between healthy and NOD samples of  $\beta$ -cells. We further show that there is a strong link between the number of granules in a sample and the healthy/NOD class it belongs to. We apply this insight to train a neural network that jointly learns both classification and counting tasks, to improve performance on classification. Further, we explore the concept of transfer learning where a network pre-trained on granule counting can be beneficial

TABLE II  
POSITIVE INFLUENCES OF  $\beta$ -CELL GRANULE COUNTING ON CLASSIFICATION FOR JOINT TRAINING (LEFT) AND TRANSFER LEARNING (RIGHT).

| GT | only           |       | GT | only           |          |
|----|----------------|-------|----|----------------|----------|
|    | classification | joint |    | classification | transfer |
| 48 | -              | 45.6  | -  | -              | -        |
| 0  | 1              | 0     | 0  | 1              | 0        |

when fine-tuning it on classification, which also yields an improved accuracy. The multi-tasking and transfer learning approaches allow for a more robust representation by including the additional counting task. This can be especially useful in the medical domain, where often only limited data is available but related additional tasks can be defined in a similar fashion.

The findings strongly support the underlying hypothesis that early onset diabetes leads to a reduction in insulin-producing granules. Nevertheless, classification on its own already delivers strong results. This indicates that there are additional factors other than the number of granules that play a major role in the decision process, such as changed appearances in the granules themselves or the surrounding tissue. Therefore, it is planned to expand this analysis to a larger scale once more data has been acquired, which can be a time-consuming process, but will allow to draw conclusions that are statistically reliable.

#### VI. ETHICAL APPROVAL

All mouse procedures were in accordance with institutional guidelines and approved by the corresponding authorities (Projects BMFW-66.010/0142-WF/V/3b/2017 for NOD/ShiLtJ mice, BMFW-66.010/0160-WF/V/3b/2017 for C57BL/6J mice).

#### ACKNOWLEDGMENT

This work was supported by the BioTechMed Graz flagship project “MIDAS”. Marc Masana acknowledges the support by the “University SAL Labs” initiative of Silicon Austria Labs (SAL).

#### REFERENCES

- [1] G. Alfian, M. Syafrudin, J. Rhee, M. Anshari, M. Mustakim, and I. Fahrurrozi, “Blood glucose prediction model for type 1 diabetes based on extreme gradient boosting,” *IOP Conference Series: Materials Science and Engineering*, vol. 803, p. 012012, 2020.
- [2] M. A. Atkinson, G. S. Eisenbarth, and A. W. Michels, “Type 1 diabetes,” *The Lancet*, vol. 383, no. 9911, pp. 69–82, 2014.
- [3] A. Burrack, T. Martinov, and B. Fife, “T cell-mediated beta cell destruction: Autoimmunity and alloimmunity in the context of type 1 diabetes,” *Frontiers in Endocrinology*, vol. 8, 2017.

- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 770–778.
- [6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [7] C. E. Mathews, S. Xue, A. Posgai, Y. L. Lightfoot, X. Li, A. Lin, C. Wasserfall, M. J. Haller, D. Schatz, and M. A. Atkinson, "Acute versus progressive onset of diabetes in nod mice: Potential implications for therapeutic interventions in type 1 diabetes," *Diabetes*, vol. 64, p. 3885–3890, 2015.
- [8] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2014.
- [9] S. J. Pan and Q. Yang, "A survey on transfer learning," *Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, 2010.
- [10] J. Pearson, F. Wong, and L. Wen, "The importance of the non obese diabetic (nod) mouse model in autoimmune diabetes," *Journal of autoimmunity*, vol. 66, pp. 76–88, 2016.
- [11] B. O. Roep, "There is something about insulin granules," *Diabetes*, vol. 69, no. 12, pp. 2575–2577, 2020.
- [12] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.
- [13] H. S. Spijker, R. B. Ravelli, A. M. Mommaas-Kienhuis, A. A. van Apeldoorn, M. A. Engelse, A. Zaldumbide, S. Bonner-Weir, T. J. Rabelink, R. C. Hoeben, H. Clevers *et al.*, "Conversion of mature human  $\beta$ -cells into glucagon-producing  $\alpha$ -cells," *Diabetes*, vol. 62, no. 7, pp. 2471–2480, 2013.
- [14] G. Tripathi and R. Kumar, "Early prediction of diabetes mellitus using machine learning," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2020, pp. 1009–1014.
- [15] B. L. Welch, "The generalisation of student's problems when several different population variances are involved," *Biometrika*, vol. 34 1-2, pp. 28–35, 1947.
- [16] A. Z. Woldaregay, E. Årsand, T. Botsis, D. Albers, L. Mamykina, and G. Hartvigsen, "Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes," *Journal of medical Internet research*, vol. 21, no. 5, p. e11030, 2019.
- [17] J. Xue, F. Min, and F. Ma, "Research on diabetes prediction method based on machine learning," *Journal of Physics: Conference Series*, vol. 1684, p. 012062, 2020.
- [18] A. Zaitcev, M. R. Eissa, Z. Hui, T. Good, J. Elliott, and M. Benaissa, "A deep neural network application for improved prediction of hba1c in type 1 diabetes," *Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2932–2941, 2020.
- [19] X. Zhang, X. Peng, and C. e. a. Han, "A unified deep-learning network to accurately segment insulin granules of different animal models imaged under different electron microscopy methodologies," *Protein Cell*, vol. 10, 2019.