3. Desenvolvimento Para o primeiro modelo, a arquitetura geral de uma rede neural convolucional foi estudada, e um modelo simplificado foi montado baseado nas redes neurais já existentes, tendo o modelo do cientista da computação Yann LeCun como base [10]. O modelo pode ser dividido em três partes conforme a Figura 1 ilustra. A primeira parte da rede neural convolucional consiste em operações de convolução, sendo a primeira camada uma função com a formatação da entrada, as especificações do filtro, e uma função de ativação “ReLU”. Em seguida, uma camada de “Pooling” foi adicionada, onde o resultado da camada de convolução é comprimido e as informações e atributos desnecessários são excluídos. Foi escolhido o tipo “Max Pooling” onde os maiores valores são preservados, e os demais descartados. Essas duas camadas se revezaram quatro vezes no modelo, sendo esse número de camadas o que apresentou o melhor resultado ao longo de várias tentativas. Na modelagem de redes neurais, a experimentação predomina, não existindo regras absolutas quanto às suas configurações. Figura 1 – Rede Neural Convolucional. O resultado é uma matriz, um mapa de atributos que foram destacados durante as repetidas operações de convolução. E para as camadas seguintes da rede neural, que são totalmente conectadas formando uma rede neural artificial para categorização na segunda parte do modelo, essas matrizes foram então transformadas em vetores com o método “Flatten”. A partir deste ponto, as camadas da rede neural artificial são completamente conectadas, tendo os vetores de atributos selecionados como entrada. Essas camadas seguintes da rede neural artificial foram criadas com o método “Dense”, significando que todos os 64 nós da camada foram conectados a todos os 64 nós da camada seguinte e assim por diante. O número de nós ou neurônios de uma camada também é experimental, geralmente um múltiplo de 2. Essas operações só são possíveis com a vetorização das matrizes convolucionais. Entre as camadas totalmente conectadas, recomendava-se o método “Dropout” para evitar o " overfitting ”, que é quando a rede neural se torna tendenciosa. Com esse método, neurônios nas camadas são ignorados aleatoriamente [11]. A função de ativação da última camada é a “Softmax”, uma função de normalização utilizada quando existem mais de duas classes ou categorias de saída, no caso do trabalho, 3 classes [12]. Por final na criação do modelo, ele é compilado. A função de compilação incluiu o otimizador Adam, um método de otimização estocástica para recalibrar os pesos dos neurônios [13]. Incluiu também a função de perda categorical\_crossentropy, [14]. E uma métrica de acurácia, que é como o modelo mede sua própria eficácia. Para o segundo modelo foi realizada a aplicação da técnica de Transferência de Aprendizado. As redes neurais pré-treinadas utilizadas nela são modelos que foram previamente treinados com base nas imagens do banco da ImageNet, que possui milhões de imagens para mais de um milhão de tipos de classificação. Com o processo de Transferência de Aprendizado, o mapeamento de atributos e a configuração dos pesos da rede neural são reaproveitados e adicionam-se as novas imagens a serem classificadas nas camadas finais de processamento para sua customização. Com isso diminui-se a necessidade de uma base de imagens muito grande e o processamento do treinamento, já que apenas os pesos das camadas adicionais precisam ser treinados [15]