**Slide 10: Dataset and Training Approach**

* **Initial Dataset:** Started with 100 individuals for each severity category.
* **Challenges:** Poor predictions after initial training.
* **Improved Approach:** Increased dataset to 400 individuals per category, keeping weight equal across all severity categories, with oversampling on a personal level (more images per person).

**Slide 11: Predictions and Observations**

* **Confusion Matrix:** To be provided later.
* **Model Performance:**
  + The first and last classes were poorly predicted.
  + The severe class was misclassified as moderate by 69%, reflecting the continuum of conditions.
* **Classification Adjustments:** Combined moderate and severe categories into a single "damage or not damage" category to improve model focus.

**Slide 6: Evaluation Metric**

* **Metric Used**: Weighted log loss and an additional metric called any\_severe\_spinal.
* **Weights by Severity Level**:
  + Normal/Mild: Weight = 1
  + Moderate: Weight = 2
  + Severe: Weight = 4
* **Submission Format**: Probability predictions for each severity level.

**Slide 7: Submission Format Example**

* **File Format**:  
  Example row: row\_id, normal\_mild, moderate, severe  
  e.g., 123456\_left\_neural\_foraminal\_narrowing\_l1\_l2, 0.333, 0.333, 0.333
* **Nulls in Rows**: Nulls are not allowed; predictions are required even for non-visible vertebrae.

**Slide 8: Model Development Strategy**

* **Model Objective**: Predict severity of spine degeneration conditions.
* **Techniques Considered**: Transfer learning with models like ResNet-50, CNN-based architectures, ensemble methods.
* **Data Augmentation and Preprocessing**: Techniques to increase model generalizability.

**Slide 9: Project Overview (Previously Slide 1)**

* **Goal**: Classify spinal conditions using MRI images.
* **Focus**: Left Neural Foraminal Narrowing and L5/S1 location for computational efficiency.
* **Method**: Leveraging Sagittal images to identify stenosis and neural foraminal narrowing.

**Slide 10: Image Categories (Previously Slide 2)**

* **Sagittal T2/STIR**
* **Sagittal T1**
* **Axial T2**

**Slide 11: Image Quality Assessment (Previously Slide 3)**

* **Key Insight**: Different image types have varying effectiveness for detecting damage.
  + **All images should be used**, but some images contribute more significantly than others.
* **Focus**: Sagittal images, as they provide a complete view of the spine, critical for condition identification.

**Slide 12: Condition Assessment (Previously Slide 4)**

* **Spinal Canal Stenosis**: Primarily assessed with **Sagittal T2/STIR**, **Sagittal T1**, and **Axial T2**.
* **Subarticular Stenosis (Left/Right)**: Best visualized with **Axial T2** and **Sagittal T2**.
* **Neural Foraminal Narrowing (Left/Right)**: Best seen with **Sagittal T2/STIR**, **Sagittal T1**, and sometimes **Axial T2**.

**Slide 13: Image Selection Rationale (Previously Slide 5)**

* **Sagittal Images**:
  + Show the entire spine, providing more context for detecting stenosis and neural foraminal narrowing.
* **Axial Images**:
  + Only show localized parts of the spine, limiting their effectiveness for detecting stenosis when only one image per person is available.
  + Risk of selecting an incorrect position if Axial images are used.

**Slide 14: Dataset and Methodology (Previously Slide 6)**

* **Dataset**: Curated from 8 global sites, spanning 5 continents.
* **Grading by Disc Levels**: Severity scores across L1/L2 to L5/S1 disc levels.
* **Goal**: Develop a standardized classification for more accurate and rapid diagnosis.

**Slide 1: Preliminary Results**

* Predictions with 100 individuals per severity classification range between **35%** and **37%** accuracy.
* Changing layers in the model did not improve the results.

**Slide 2: New Sample Strategy**

* Cases with **medium** and **severe** severity are **undersampled** in terms of persons, but **balanced** in terms of images.
* For each person, there are **4000 images**.

**Slide 3: Training Details**

* Train-test split is set at **80%** for training and **20%** for testing.
* Calculation time on the training set is approximately **4 hours**.

**Slide 4: Early Stopping**

* **Early stopping** was applied after **6 epochs** due to divergence between the **validation** and **training datasets**.

**Slide 1: Prediction Challenges for the First Category**

* **First Category Prediction Issue**: The **first category** (Normal/Mild) is predicted very poorly.
* **Misclassification**: Frequently predicted as the **second category** (Moderate).
* **Spinal Damage Issue**: There is **no clear distinction** between medium and severe severity.
* **Subjectivity in Diagnosis**: Doctors often decide how to classify conditions, leading to variations in interpretation.

**Slide 2: Combining Medium and Severe Categories**

* **Decision**: Due to the lack of clear distinction, the **Medium** and **Severe** categories have been combined into a **single "Damage or Not Damage" classification**.
* **Rationale**:
  + Some spinal conditions involve prolapse or protrusion, while others do not.
  + By combining these categories, a more practical and consistent classification approach is adopted.

 The **training data** contains images labeled with **damage severity** (e.g., mild, moderate, severe).

*  However, the **decision of how much damage qualifies** as "big" or "small" is **not standardized**.