Table D.8: Annotations used in the algorithms. Note that "disease names" can represent both unnormalized disease names or standard disease names.

Descriptions	Notations
Axis word	$a1$, $a2$, $a3$, etc.
List of axis words	$A1$, $A2$, $A3$, etc.
Axis type - Disease Center	dce
Axis type - Anatomical Region	al
Axis type - Disease Characteristic	dch
Larger region	lar
List of shared axis words between two diseases	SA
List of differing axis words between two diseases	DiA
List of differing axis words in the first disease when comparing two diseases	$DiA1$
List of differing axis words in the second disease when comparing two diseases	$DiA2$
Unnormalized disease names (UDN)	$u1$, $u2$, $u3$, etc.
Standard disease names (SDN)	$s1$, $s2$, $s3$, etc.
Disease names (can be either a UDN or an SDN)	$d1$, $d2$, $d3$, etc.

Algorithm 1 Axis-word Replacement 1 (AR1)

```
 1: Input:
        $training\_set$ - List of disease pairs from the disease name
         normalization training set.
        $ICD\_list$ - The standard ICD system.
 2: Output: $augmented\_pairs$ - List of augmented disease pairs.
 3: procedure AR1($training\_set, ICD\_list$)
 4:     $augmented\_pairs \leftarrow []$
 5:     for each $d1$ in $(training\_set \cup ICD\_list)$ do
 6:         $A1 \leftarrow NER(d1)$
 7:         for each $s1$ in $ICD\_list$ do
 8:             $A2 \leftarrow NER(s1)$
 9:             $SA, DiA1, DiA2 \leftarrow comparing\_axis\_words(A1, A2)$
10:             if len($SA$) $\neq 0$ and len($DiA1$) = len($DiA2$) = 1 then
11:                 $d2 \leftarrow d1.replace\_axis(DiA1[0], DiA2[0])$
12:                 $augmented\_pairs.append((d2, s1))$
13:             end if
14:         end for
15:     end for
16:     return $augmented\_pairs$
17: end procedure
```

Algorithm 2 Axis-word Replacement 2 (AR2)

```
 1: Input:
        $training\_set$ - List of disease pairs from the disease name
         normalization training set.
        $ICD\_list$ - The standard ICD system.
 2: Output: $augmented\_pairs$ - List of augmented disease pairs.
 3: procedure AR2($training\_set, ICD\_list$)
 4:     $augmented\_pairs \leftarrow []$
 5:     for each $(u1, s1)$ in $training\_set$ do
 6:         $A1 \leftarrow NER(u1)$
 7:         $A2 \leftarrow NER(s1)$
 8:         if $A1 = A2$ then
 9:             for each $s2$ in $ICD\_list$ do
10:                 $A3 \leftarrow NER(s2)$
11:                 $SA, DiA1, DiA2 \leftarrow comparing\_axis\_words(A2, A3)$
12:                 if len($A2$) =len($A3$) and len($DiA1$)=len($DiA2$) $= 1$ then
13:                     $s3 \leftarrow s1.replace\_axis(DiA1[0], DiA2[0])$
14:                     $u2 \leftarrow u1.replace\_axis(DiA1[0], DiA2[0])$
15:                     $augmented\_pairs.append((u2, s3))$
16:                 end if
17:             end for
18:         end if
19:     end for
20:     return $augmented\_pairs$
21: end procedure
```

Algorithm 3 Multi-Granularity Aggregation - Code (MGA-Code)

```
1: **Input:**
       $training\_set$ - List of disease pairs from the disease name
        normalization training set.
       $ICD\_list$ - The standard ICD system.
2: **Output:** $augmented\_pairs$ - List of augmented disease pairs.
3: procedure MGA-CODE($training\_set, ICD\_list$)
4:     $augmented\_pairs \leftarrow []$
5:     for each $d1$ in ($training\_set \cup ICD\_list$) do
6:         if $len$(ICD-code($d1$)) = 6 then
7:             $code_6$ =ICD-code($d1$)
8:             $code_4 = code_6[0:3]$ // Extract the first four digits
9:             $s1$ =map_disease($code_4$)
10:            $augmented\_pairs.append((d1, s1))$
11:        end if
12:    end for
13:    return $augmented\_pairs$
14: end procedure
```

Algorithm 4 Multi-Granularity Aggregation - Region (MGA-Region)

```
1: Input:
       $training\_set$ - List of disease pairs from the disease name
        normalization training set.
       $ICD\_list$ - The standard ICD system.
2: Output: $augmented\_pairs$ - List of augmented disease pairs.
3: procedure MGA-Region($training\_set, ICD\_list$)
4:     $augmented\_pairs \leftarrow []$
5:     for each $d1$ in ($training\_set \cup ICD\_list$) do
6:         $A1 \leftarrow NER(d1)$
7:         for each $s1$ in $ICD\_list$ do
8:             $A2 \leftarrow NER(s1)$
9:             $SA, DiA1, DiA2 \leftarrow comparing\_axis\_words(A1, A2)$
10:            if len($SA$) $\geq 1$ and len($DiA1$)=len($DiA2$) $= 1$ then
11:                if $type(DiA1) = al$ and $DiA2[0]$=$DiA1[0].lar$ then
12:                    $augmented\_pairs.append((d1, s1))$
13:                end if
14:            end if
15:        end for
16:    end for
17:    return $augmented\_pairs$
18: end procedure
```