Supplementary Material for CliffPhys: Camera-based Respiratory Measurement using Clifford Neural Networks

Omar Ghezzi[1], Giuseppe Boccignone[1], Giuliano Grossi[1], Raffaella Lanzarotti[1], and Alessandro D'Amelio[1]

PHuSe Lab - Università degli Studi di Milano
omar.ghezzi@studenti.unimi.it
{giuseppe.boccignone,giuliano.grossi,
raffaella.lanzarotti,alessandro.damelio}@unimi.it

1 Mathematical Background

1.1 Real Clifford Algebra $Cl_{p,q}$

The Clifford algebra [2, 5, 6] is a specific algebraic structure, known as algebra over a field, which unifies heterogeneous substructures into a single environment. This algebra blurs the distinction between vector spaces and numerical fields, allowing the representation and multiplication of elements belonging to both domains. Formally, an algebra V over a Field K is a vector space $(V, +_V, \cdot_{KV})$, implicitly defined on the field $(K, +_K, \cdot_K)$ of scalars, together with a bilinear composition law $\cdot_V : V \times V \to V$.

The Clifford Algebra over the real field $\mathbb{R}$ enriches $\mathbb{R}^n$ with a specific bilinear operator known as the Clifford product. Consider the vector space $(\mathbb{R}^n, +_{\mathbb{R}^n}, \cdot_{\mathbb{R}\mathbb{R}^n})$ over the real field $(\mathbb{R}, +_{\mathbb{R}}, \cdot_{\mathbb{R}})$. Endow $\mathbb{R}^n$ with a symmetric endomorphism $f_A(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$, represented by a matrix $A \in \mathbb{R}^{n\times n} : A^T = A$, s.t. the associated mapping $q_A(\cdot)$ is a non-degenerate quadratic form:

$$q_A(\cdot) : \mathbb{R}^n \to \mathbb{R} \tag{1}$$

$$\forall \mathbf{x} \in \mathbb{R}^n, \;\; \mathbf{x} \mapsto q_A(\boldsymbol{x}) = \mathbf{x}^T A \mathbf{x} \;\; . \tag{2}$$

Any symmetric matrix A can be orthogonally diagonalized, resulting in a diagonal matrix Λ with eigenvalues $\lambda_i \in \mathbb{R}$ and eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_n$, forming an orthonormal basis for $\mathbb{R}^n$. This results in diagonalizing also $q_A(\cdot)$, meaning that $q_A(\mathbf{v}_i) = q_\Lambda(\mathbf{e}_i) = \lambda_i$:

$$q_A(\mathbf{v}_i) = \mathbf{v}_i^T A \mathbf{v}_i = (S\mathbf{e}_i)^T A S \mathbf{e}_i = \mathbf{e}_i^T S^T A S \mathbf{e}_i = \mathbf{e}_i^T \Lambda \mathbf{e}_i = \lambda_i \;\; , \tag{3}$$

with $q_\Lambda(\cdot)$ being the diagonalized quadratic form, $S = [\mathbf{v}_1|...|\mathbf{v}_n] \in \mathbb{R}^{n\times n}$ being the orthogonal change of basis matrix and $\mathbf{e}_i$ being the coordinates of $\mathbf{v}_i$ w.r.t. the orthonormal basis composed by $\mathbf{v}_1, ..., \mathbf{v}_n$. Let now reorganize the vectors $\mathbf{v}_i$ such that the first p correspond to positive eigenvalues, while the remaining q correspond to negative eigenvalues. Following this sorting, a change of basis

is considered from the orthonormal basis comprised of $\mathbf{v}_1, ..., \mathbf{v}_n$ to a new basis consisting of vectors $\mathbf{u}_1, ..., \mathbf{u}_n$. These new basis elements are explicitly defined as follows: $\forall i = 1, ..., p, : \mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{v}_i$ and $\forall i = p+1, ..., q, : \mathbf{u}_i = \frac{1}{\sqrt{-\lambda_i}}\mathbf{v}_i$. The current change of coordinate system achieves the normalization of the quadratic form $q_A(\cdot)$, guaranteeing the following condition:

$$q_A(\mathbf{u}_i) = q_G(\mathbf{e}_i) = \gamma_i = \begin{cases} +1 & 1 \leq i \leq p \\ -1 & p < i \leq p+q \end{cases} \quad , \tag{4}$$

in the case $1 \leq i \leq p$, Eq. (4) is obtained as follows:

$$\begin{aligned} q_A(\mathbf{u}_i) &= \mathbf{u}_i^T A \mathbf{u}_i && (5)\\ &= (\frac{1}{\sqrt{\lambda_i}}\mathbf{v}_i)^T A \mathbf{v}_i \frac{1}{\sqrt{\lambda_i}} && (6)\\ &= \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i^T S^T A S \mathbf{e}_i \frac{1}{\sqrt{\lambda_i}} && (7)\\ &= \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i^T \Lambda \mathbf{e}_i \frac{1}{\sqrt{\lambda_i}} && (8)\\ &= \mathbf{e}_i^T G \mathbf{e}_i && (9)\\ &= \gamma_i = 1 \quad . && (10) \end{aligned}$$

The Clifford algebra $(Cl_{p,q}, +_{Cl_{p,q}}, \cdot_{Cl_{p,q}})$ with signature (p, q) is a real associative algebra of dimension 2^n (where $n = p+q$). It is defined by requiring that, for each canonical vector $\mathbf{e}_1, ..., \mathbf{e}_n \in \mathbb{R}^n$, represented[1] by $e_1, ..., e_n \in Cl_{p,q}$, the Clifford product[2] $\cdot_{Cl_{p,q}}$ between e_i and itself equals $q_A(\mathbf{u}_i)$ (or $q_G(\mathbf{e}_i)$), while it anti-commutes for e_i, e_j, $i \neq j$:

$$\begin{aligned} e_i e_i = e_i^2 = \gamma_i, \quad \forall i = 1, ..., p+q \quad , && (11)\\ e_i e_j = -e_j e_i \quad \forall i < j \quad . && (12) \end{aligned}$$

The Clifford product $\cdot_{Cl_{p,q}}$ is associative, distributive (with respect to the sum of elements within the algebra) and preserves the norm of vectors in the space $\mathbb{R}^n$ in which it is defined. Specifically, starting to build $Cl_{2,0}$ from $\mathbb{R}^2$, by taking $p = 2, q = 0$ (i.e. $\gamma_1 = \gamma_2 = 1$), the Clifford product, between any vector $\mathbf{a} \in \mathbb{R}^2$ (represented by $a \in Cl_{2,0}$) and itself, equals the squared norm $||\mathbf{a}||^2_{\mathbb{R}^2} \in \mathbb{R}$:

$$a^2 = ||\mathbf{a}||^2_{\mathbb{R}^2} \quad . \tag{13}$$

[1] To ensure consistent notation, we introduce three symbols for denoting the i-th basis element in $\mathbb{R}^n$. For instance, in $\mathbb{R}^2$ and $Cl_{2,0}$, $\mathbf{e}_1 = (1, 0)^T$ represents a vector in $\mathbb{R}^2$, $e_1 = 0 + 1e_1 + 0e_2 + 0e_{12}$ represents an element of $Cl_{2,0}$, and $\underline{e}_1 = (0, 1, 0, 0)^T$ denotes the coefficients of e_1, i.e a vector in $\mathbb{R}^{2^{p+q}}$, where $p = 2, q = 0$.

[2] The Clifford product, denoted as $\cdot_{Cl_{p,q}}$, is distinguished from other products by this notation. When context permits, it's abbreviated to ab to denote a Clifford product between two elements $a, b \in Cl_{p,q}$.

This is motivated by the fact that it must be $a^2 = q_G(\mathbf{a})$ and, in this case, the metric G reduces to the identity matrix $I \in \mathbb{R}^{2\times 2}$:

$$\begin{aligned} q_G(\mathbf{a}) &= \mathbf{a}^T G \mathbf{a} && (14)\\ &= (\mathrm{a}_1 \mathrm{a}_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{a}_1 \\ \mathrm{a}_2 \end{pmatrix} && (15)\\ &= \mathrm{a}_1^2 + \mathrm{a}_2^2 && (16)\\ &= ||\mathbf{a}||^2_{\mathbb{R}^2} \quad . && (17) \end{aligned}$$

Explicitly computing $a^2 = ||\mathbf{a}||^2_{\mathbb{R}^2}$, we can observe how the multiplicative rules of the Clifford product, for $Cl_{2,0}$, preserve the equality:

$$\begin{aligned} ||\mathbf{a}||^2_{\mathbb{R}^2} &= a^2 && (18)\\ \mathrm{a}_1^2 + \mathrm{a}_2^2 &= (a_1 e_1 + a_2 e_2)^2 && (19)\\ \mathrm{a}_1^2 + \mathrm{a}_2^2 &= a_1^2 e_1^2 + a_2^2 e_2^2 + a_1 a_2 (e_1 e_2 + e_2 e_1) \quad . && (20) \end{aligned}$$

The only way to satisfy this equation (recalling that $a_i = \mathrm{a}_i$, $i = 1, 2$, by definition) is to set:

$$e_1^2 = e_2^2 = 1 \quad , \tag{21}$$

$$e_1 e_2 + e_2 e_1 = 0 \quad \Rightarrow \quad e_1 e_2 = -e_2 e_1, \qquad \text{with:} \quad e_1 e_2 = e_{12} \qquad , \tag{22}$$

which are exactly the Clifford multiplication rules for $Cl_{2,0}$.

1.2 Clifford product of vectors: blades and multivectors

The specific multiplicative rules of the Clifford product in $Cl_{2,0}$ are essential constraints ensuring that the Clifford product of an element by itself equates its squared norm, when such an element represents a vector in $\mathbb{R}^2$. We now illustrate the algebraic form of the Clifford product between two vectors in $\mathbb{R}^n$, exploiting the product in $\mathbb{R}^2$ as an example. Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ (represented by $a, b \in Cl_{2,0}$), the Clifford product $a \cdot_{Cl_{2,0}} b$ is given by:

$$\begin{aligned} ab &= (a_1 e_1 + a_2 e_2)(b_1 e_1 + b_2 e_2) && (23)\\ &= a_1 b_1 e_1^2 + a_2 b_2 e_2^2 + a_1 b_2 e_{12} + a_2 b_1 e_{21} && (24)\\ &= (a_1 b_1 + a_2 b_2) + (a_1 b_2 - a_2 b_1) e_{12} && (25)\\ &= \langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^2} + \mathbf{a} \wedge_{\mathbb{R}^2} \mathbf{b} \quad . && (26) \end{aligned}$$

Specifically, the Clifford product between two vectors in $\mathbb{R}^2$ results in a complex number. This occurs because the scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}^2}$ yields a scalar part, while the wedge product $\wedge_{\mathbb{R}^2}$ between vectors yields an immaginary part, i.e. an element of the linear bivector space $\bigwedge^2 \mathbb{R}^2$. This space is spanned by e_{12}, which satisfies $e_{12}^2 = -1$.

More broadly, for two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$, represented by elements a, b in $Cl_{p,q}$, their geometric product is defined as:

$$ab = \langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^n} + \mathbf{a} \wedge_{\mathbb{R}^n} \mathbf{b} \quad , \tag{27}$$

such that the inner product, $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^n}$, is its symmetric part and the exterior (or wedge) product, $\mathbf{a} \wedge \mathbf{b}_{\mathbb{R}^n}$, its antisymmetric part. As the inner product between pairs of distinct basis elements vanishes (i.e. $\langle \mathbf{e}_i, \mathbf{e}_j \rangle_{\mathbb{R}^n} = 0$ for $i \neq j$), the bivector e_{ij} aligns with the exterior product $\mathbf{e}_i \wedge \mathbf{e}_j$. It is possible to capture the terms on the right hand side of Eq. (27) in the following way [5]:

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^n} = \frac{1}{2} (ab + ba) \quad , \tag{28}$$

$$\mathbf{a} \wedge_{\mathbb{R}^n} \mathbf{b} = \frac{1}{2} (ab - ba) \quad . \tag{29}$$

More in general, the Clifford product between k nonzero orthogonal vectors in $\mathbb{R}^n$ is referred to as a k-blade, or simply a blade of grade k. As the k considered vectors are orthogonal, a k-blade aligns with the wedge product of these vectors, yielding an element of the exterior algebra $\bigwedge_{\mathbb{R}^n}$ of the linear space $\mathbb{R}^n$. This element represents the subspace within $\mathbb{R}^n$ that is spanned by the aforementioned orthogonal vectors. Specifically, considering π as a permutation of k arbitrary integers in the range $\{1, \ldots, n\}$ and $\pi(i)$ as the i-th element in π, we are interested in k-blades as multiples of the basis element $\mathbf{e}_{\pi(1)} \wedge \cdots \wedge \mathbf{e}_{\pi(k)}$, which generates one of the $\binom{n}{k}$ subspaces of $\bigwedge^k \mathbb{R}^n \subset \bigwedge \mathbb{R}^n$:

$$\boldsymbol{X}_\pi = x_{\pi(1)} \mathbf{e}_{\pi(1)} \wedge \cdots \wedge x_{\pi(k)} \mathbf{e}_{\pi(k)} \quad . \tag{30}$$

Eq. (30) reduces to the Clifford product $x_{\pi(1)} e_{\pi(1)} \cdots x_{\pi(k)} e_{\pi(k)} \in Cl_{p,q}$. The significance of the k-blade $\boldsymbol{X}_\pi$ lies in the fact that it represents, simultaneously, the subspace of $\mathbb{R}^n$ spanned by $\mathbf{e}_{\pi(1)}, \ldots, \mathbf{e}_{\pi(k)}$ and an oriented hypervolume element with magnitude $|\boldsymbol{X}_\pi| = x_{\pi(1)} \cdots x_{\pi(k)}$ [6].

The Clifford algebra $Cl_{p,q}$ has a multivector structure, meaning that a generic element $x \in Cl_{p,q}$ can be viewed as a linear combinations of blades of the form reported in Eq. (30). For example, given an element $x \in Cl_{3,0}$, such that:

$$x = x_0 + x_1 e_1 + x_{23} e_{23} + x_{123} e_{123} \quad \in Cl_{3,0} \qquad , \tag{31}$$

it is possible to view x as composed by the following blades:

$$\boldsymbol{X}_0 = x_0 \qquad \text{(zero-graded blades are non-zero scalars)} \ , \tag{32}$$

$$\boldsymbol{X}_1 = x_1 \mathbf{e}_1, \qquad \text{with } \pi = 1 \ , \tag{33}$$

$$\boldsymbol{X}_{23} = x_2 \mathbf{e}_2 \wedge x_3 \mathbf{e}_3, \qquad \text{with } \pi = 2, 3 \ , \tag{34}$$

$$\boldsymbol{X}_{123} = x_1 \mathbf{e}_1 \wedge x_2 \mathbf{e}_2 \wedge x_3 \mathbf{e}_3, \qquad \text{with } \pi = 1, 2, 3 \ . \tag{35}$$

In section 1.4 we will highlight that algebraic operations on blades (and, by linearity of the Clifford product, also on multivectors) correspond to geometric operations on the subspaces they represent. In the next subsection, we list the main algebras of interest in this work.

1.3 Examples of Clifford Algebras

Clifford Algebra $Cl_{2,0}$ The Clifford algebra $Cl_{2,0}$ is a 4-dimensional real algebra with basis elements $\{1, e_1, e_2, e_{12}\}$, s.t. $\forall a \in Cl_{2,0}$, $a = a_0 + a_1 e_1 + a_2 e_2 + a_{12} e_{12}$, and multiplication rules:

$$e_i^2 = 1 \qquad \forall i = 1, 2, 3 \quad , \tag{36}$$

$$e_i e_i = -e_j e_i \qquad \forall i = 1, 2, 3,\ i \neq j \quad . \tag{37}$$

For $Cl_{2,0}$ the following isomorphism holds: $Cl_{2,0} \simeq \mathbb{R} \oplus \mathbb{R}^2 \oplus \bigwedge^2 \mathbb{R}^2$, with:

$$\mathbb{R} \simeq \{x + 0e_1 + 0e_2 + 0e_{12} | x \in \mathbb{R}\} \subset Cl_{2,0} \quad , \tag{38}$$

$$\mathbb{R}^2 \simeq \{0 + xe_1 + ye_2 + 0e_{12} | (x, y) \in \mathbb{R}^2\} \subset Cl_{2,0} \quad , \tag{39}$$

$$\mathbb{C} \simeq \{x + 0e_1 + 0e_2 + ye_{12} | x + iy \in \mathbb{C}\} \subset Cl_{2,0} \quad , \tag{40}$$

$$\bigwedge^2 \mathbb{R}^2 \simeq \{0 + 0e_1 + 0e_2 + xe_{12} | x \in \mathbb{R}\} \subset Cl_{2,0} \quad . \tag{41}$$

Clifford Algebra $Cl_{0,2}$ The Clifford algebra $Cl_{0,2}$ is a 4-dimensional real algebra with basis elements $\{1, e_1, e_2, e_{12}\}$, s.t. $\forall a \in Cl_{0,2}$, $a = a_0 + a_1 e_1 + a_2 e_2 + a_{12} e_{12}$, and multiplication rules:

$$e_i^2 = -1 \qquad \forall i = 1, 2, 3 \quad , \tag{42}$$

$$e_i e_i = -e_j e_i \qquad \forall i = 1, 2, 3,\ i \neq j \quad . \tag{43}$$

For $Cl_{0,2}$ the following isomorphism holds: $Cl_{0,2} \simeq \mathbb{H}$, with $(\mathbb{H}, +_{\mathbb{H}}, \cdot_{\mathbb{H}})$ denoting the real associative algebra of quaternions. Quaternions are generalized complex numbers of the form:

$$\mathbb{H} = \{q = q_0 + q_1 i + q_2 j + q_3 k \,|\, i^2 = j^2 = k^2 = ijk = -1\} \quad , \tag{44}$$

A quaternion is a sum of a scalar and a vector, which are called respectively real part, $\mathbf{Re}(\mathbb{H}) = \{q = q_0 + 0i + 0j + 0k \mid q_0 \in \mathbb{R}\} \simeq \mathbb{R}$, and pure part, $\mathbf{Pu}(\mathbb{H}) = \{q = 0 + q_1 i + q_2 j + q_3 k \,|\, \mathbf{q} = (q_1, q_2, q_3) \in \mathbb{R}^3\} \simeq \mathbb{R}^3$. The isomorphism $Cl_{0,2} \simeq \mathbb{H}$ is obtained setting $e_1 = i$, $e_2 = j$, and $e_{12} = k$.

Clifford Algebra $Cl_{3,0}$ The Clifford algebra $Cl_{3,0}$ is a 8-dimensional real algebra with basis elements $\{1, e_1, e_2, e_3, e_{12}, e_{23}, e_{13} e_{123}\}$, s.t. $\forall a \in Cl_{3,0}$, $a = a_0 + a_1 e_1 + a_2 e_2 + a_3 e_3 + a_{12} e_{12} + a_{23} e_{23} + a_{13} e_{13} + a_{123} e_{123}$, and multiplication rules:

$$e_i^2 = 1 \qquad \forall i = 1, 2, 3 \quad , \tag{45}$$

$$e_i e_i = -e_j e_i \qquad \forall i = 1, 2, 3,\ i \neq j \quad . \tag{46}$$

For $Cl_{3,0}$ the following isomorphism holds: $Cl_{3,0} \simeq \mathbb{R} \oplus \mathbb{R}^3 \oplus \bigwedge^2 \mathbb{R}^3 \oplus \bigwedge^3 \mathbb{R}^3$, with:

$$\bigwedge^2 \mathbb{R}^3 \oplus \bigwedge^3 \mathbb{R}^3 \simeq \mathbb{H} \subset Cl_{3,0} \quad , \tag{47}$$

$$\mathbb{R} \oplus \bigwedge^3 \mathbb{R}^3 \simeq \mathbb{C} \subset Cl_{3,0} \quad . \tag{48}$$

Clifford Algebra $Cl_{0,3}$ The Clifford algebra $Cl_{0,3}$ is a 8-dimensional real algebra with basis elements $\{1, e_1, e_2, e_3, e_{12}, e_{23}, e_{13}e_{123}\}$, s.t. $\forall a \in Cl_{0,3}$, $a = a_0 + a_1e_1 + a_2e_2 + a_3e_3 + a_{12}e_{12} + a_{23}e_{23} + a_{13}e_{13} + a_{123}e_{123}$, and multiplication rules:

$$e_i^2 = -1 \qquad \forall i = 1, 2, 3 \quad , \tag{49}$$

$$e_ie_i = -e_je_i \qquad \forall i = 1, 2, 3,\ i \neq j \quad . \tag{50}$$

For $Cl_{0,3}$ the following isomorphism holds: $Cl_{0,3} \simeq \mathbb{H} \oplus \mathbb{H}$.

1.4 Orthogonal transformations in $Cl_{p,q}$

In the algebraic framework presented so far, the geometric inductive bias can be associated with Clifford neural layers through orthogonal transformations given by an even number of reflections. Given two vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$, represented by $v, u \in Cl_{p,q}$, we can reflect $\mathbf{v}$ through the hyperplane with normal $\mathbf{u}$ by sandwiching v between u and u^{-1}, or equivalently,

$$v \mapsto -uvu^{-1} \quad . \tag{51}$$

Let $\boldsymbol{X}_{1\ldots k}$ denotes the k-blade in $\bigwedge^k \mathbb{R}^k$ formed by taking the wedge product of the k basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$ of the subspace $\mathbb{R}^k \subseteq \mathbb{R}^n$:

$$\boldsymbol{X}_{1\ldots k} = x_1\mathbf{e}_1 \wedge \cdots \wedge x_k\mathbf{e}_k \quad . \tag{52}$$

Given a permutation $\pi = \pi(1), \ldots, \pi(k)$ of the indices in $S = \{1, ..., k\}$, $\boldsymbol{X}_\pi$ denotes the k-blade formed by taking the wedge product of the basis vectors $\mathbf{e}_{\pi(1)}, \ldots, \mathbf{e}_{\pi(k)}$[3]:

$$\boldsymbol{X}_\pi = x_{\pi(1)}\mathbf{e}_{\pi(1)} \wedge \cdots \wedge x_{\pi(k)}\mathbf{e}_{\pi(k)} \quad . \tag{53}$$

Choose an arbitrary element $\pi(i)$, which occupies the i-th position in π. The k-blade $\boldsymbol{X}_\pi$ can be recursively written as follows[4]:

$$\boldsymbol{X}_\pi = (-1)^{k-i}\boldsymbol{X}_{\pi\setminus\pi(i)} \wedge \boldsymbol{X}_{\pi(i)} \quad , \tag{58}$$

[3] For instance, let's consider the case where $k = 3$, and therefore $S = \{1, 2, 3\}$. Choosing $\pi = 1, 2, 3$ we have $\boldsymbol{X}_{123} = x_1\mathbf{e}_1 \wedge x_2\mathbf{e}_2 \wedge x_3\mathbf{e}_3$. Conversely, with $\pi = 2, 3, 1$, we have $\boldsymbol{X}_{231} = x_2\mathbf{e}_2 \wedge x_3\mathbf{e}_3 \wedge x_1\mathbf{e}_1$.

[4] For instance, if $k = 3$ and $\pi = 2, 3, 1$, choose $\pi(2) = 3$ (the element of π in position $i = 2$), then:

$$\boldsymbol{X}_{231} = (-1)^{3-2}\boldsymbol{X}_{21} \wedge \boldsymbol{X}_3 \tag{54}$$

$$= -x_2\mathbf{e}_2 \wedge x_1\mathbf{e}_1 \wedge x_3\mathbf{e}_3 \tag{55}$$

$$= -(-x_2\mathbf{e}_2 \wedge x_3\mathbf{e}_3 \wedge x_1\mathbf{e}_1) \tag{56}$$

$$= x_2\mathbf{e}_2 \wedge x_3\mathbf{e}_3 \wedge x_1\mathbf{e}_1 \tag{57}$$

In the equation, the right-hand side indicates that the i-th term $\boldsymbol{X}_{\pi(i)} = x_{\pi(i)}\mathbf{e}_{\pi(i)}$ has been moved to the end of the product, requiring $k-i$ right swaps of an equal number of neighboring terms. At each step of the recursion, the following updates are performed: $S = S\backslash\{\pi(i)\}$, $\pi = \pi\backslash\pi(i)$, $k = k-1$ and a new element $\pi(i) \in S$ is selected, until k reaches 1. The 0-blade is equivalent to the neutral element concerning the wedge product (i.e., $\boldsymbol{X}_\emptyset = \mathbf{1}$). Consequently, Eq. (58) becomes $\boldsymbol{X}_p = \boldsymbol{X}_\emptyset \wedge \boldsymbol{X}_p$, for $k = 1$.

By writing this way, we can prove by induction that equation Eq. (51) extends to cases where the element multiplied in the sandwich represents a k-blade $\boldsymbol{X}_\pi$, rather than a vector. We can reflect[5] $\boldsymbol{X}_\pi \in \bigwedge^k \mathbb{R}^k$, for any choosen k, π and $\pi(i)$, through the hyperplane with $\mathbf{u}$ as its normal vector, by sandwiching $\boldsymbol{X}_\pi$ between $-\mathbf{u}$ and $\mathbf{u}^{-1}$:

$$\boldsymbol{X}_\pi \mapsto -\mathbf{u}\boldsymbol{X}_\pi\mathbf{u}^{-1} \quad . \tag{59}$$

For $k = 1$ the proof is obvious, let us prove it for $k > 1$:

$$-\mathbf{u}\boldsymbol{X}_\pi\mathbf{u}^{-1} = -1(-1)^{k-i}\ \left(\mathbf{u}\boldsymbol{X}_{\pi\backslash\pi(i)}\mathbf{u}^{-1} \wedge \mathbf{u}\boldsymbol{X}_{\pi(i)}\mathbf{u}^{-1}\right) \tag{60}$$

$$= -\frac{1}{2}(-1)^{k-i}\ \left(\mathbf{u}\boldsymbol{X}_{\pi\backslash\pi(i)}\mathbf{u}^{-1}\mathbf{u}\boldsymbol{X}_{\pi(i)}\mathbf{u}^{-1} - \mathbf{u}\boldsymbol{X}_{\pi(i)}\mathbf{u}^{-1}\mathbf{u}\boldsymbol{X}_{\pi\backslash\pi(i)}\mathbf{u}^{-1}\right) \tag{61}$$

$$= -\frac{1}{2}(-1)^{k-i}\ \left(\mathbf{u}\left(\boldsymbol{X}_{\pi\backslash\pi(i)}\boldsymbol{X}_{\pi(i)}\right)\mathbf{u}^{-1} - \mathbf{u}\left(\boldsymbol{X}_{\pi(i)}\boldsymbol{X}_{\pi\backslash\pi(i)}\right)\mathbf{u}^{-1}\right) \tag{62}$$

$$= -\mathbf{u}\left(\frac{1}{2}(-1)^{k-i}\boldsymbol{X}_{\pi\backslash\pi(i)}\boldsymbol{X}_{\pi(i)} - \boldsymbol{X}_{\pi(i)}\boldsymbol{X}_{\pi\backslash\pi(i)}\right)\mathbf{u}^{-1} \tag{63}$$

$$= -\mathbf{u}\left((-1)^{k-i}\boldsymbol{X}_{\pi\backslash\pi(i)} \wedge \boldsymbol{X}_{\pi(i)}\right)\mathbf{u}^{-1} \quad , \tag{64}$$

where we use the distributivity of $\cdot_{Cl_{p,q}}$ and, in the step Eq. (61), the equivalence Eq. (29) which express the wedge product in terms of the Clifford product between elements in $Cl_{p,q}$.

Since the Clifford product $\cdot_{Cl_{p,q}}$ is a linear operator, the result in Eq. (59) can be also extended to multivectors in $Cl_{p,q}$:

$$x \mapsto -uxu^{-1} \quad . \tag{65}$$

Racalling the specific multivector x proposed as an example in Eq. (31), reflecting x through the hyperplane with norm u coincides with the reflection of each of its composing blades:

$$-u\left(x_0 + x_1e_1 + x_{23}e_{23} + x_{123}e_{123}\right)u^{-1} = \tag{66}$$

$$= -u\,x_0\,u^{-1} - u\,x_1e_1\,u^{-1} - u\,x_{23}e_{23}\,u^{-1} - u\,x_{123}e_{123}\,u^{-1} \tag{67}$$

$$= \boldsymbol{X}_0 - \mathbf{u}\boldsymbol{X}_1\mathbf{u}^{-1} - \mathbf{u}\boldsymbol{X}_{23}\mathbf{u}^{-1} - \mathbf{u}\boldsymbol{X}_{123}\mathbf{u}^{-1} \quad . \tag{68}$$

[5] The definition in Eq. (59) and the proof in Eq. (60) should involve two elements $r, u \in Cl_{p,q}$, representing $\boldsymbol{X}_\pi \in \bigwedge^k \mathbb{R}^k$ and $\mathbf{u} \in \mathbb{R}^n$, respectively. In this cases, to make the notation a bit more intuitive, we interchangeably refer to $\boldsymbol{X}_\pi$ as r and $\mathbf{u}$ as u.

By the Cartan-Dieudonné theorem, higher-order orthogonal transformations in an n-dimensional space can be decomposed into $m \leq n$ reflections. Therefore, given the geometric product $u_1 \cdots u_m$ of m vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^n$, any orthogonal transformation of a generic multivector $x \in Cl_{p,q}$ takes the form:

$$x \mapsto (-1)^m \, u_1 \cdots u_m \, x \, (u_1 \cdots u_m)^{-1} \quad . \tag{69}$$

This means that the orthogonal transformation of a multivector in $Cl_{p,q}$ is obtained by transforming (rotating, reflecting etc.) each of its blades, or equivalently, each of the subspaces they represents.

1.5 Clifford product as a matrix multiplication

The Clifford product $\cdot_{Cl_{p,q}}$ can be equivalently represented in matrix form. Specifically, for any pair of elements in a four-dimensional Clifford algebra $a, b \in Cl_{p,q}$, $2^{p+q} = 4$, the coefficients of the left product ab and the right product ba, between a and b, take on the following matrix forms:

$$W_a^L \, \underline{b} = \begin{pmatrix} a_0 & \gamma_1 a_1 & \gamma_2 a_2 & -\gamma_1\gamma_2 a_{12} \\ a_1 & a_0 & \gamma_2 a_{12} & -\gamma_2 a_2 \\ a_2 & -\gamma_1 a_{12} & a_0 & \gamma_1 a_1 \\ a_{12} & -a_2 & a_1 & a_0 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_{12} \end{pmatrix} \quad , \tag{70}$$

$$W_a^R \, \underline{b} = \begin{pmatrix} a_0 & \gamma_1 a_1 & \gamma_2 a_2 & -\gamma_1\gamma_2 a_{12} \\ a_1 & a_0 & -\gamma_2 a_{12} & \gamma_2 a_2 \\ a_2 & \gamma_1 a_{12} & a_0 & -\gamma_1 a_1 \\ a_{12} & a_2 & -a_1 & a_0 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_{12} \end{pmatrix} \quad , \tag{71}$$

with $b \in Cl_{p,q}, p+q = 4$ being represented by $\underline{b} \in \mathbb{R}^4$, using the isomorphism $Cl_{p,q} \simeq \mathbb{R}^{2^{pq}}$. The product aba can be formalized as follows:

$$aba \mapsto W_{aa}^{RL} \underline{b} \quad . \tag{72}$$

This can be proved by letting $c = ab$, therefore $c = ab \mapsto \underline{c} = W_a^L \underline{b}$, while $ca \mapsto W_a^R \underline{c} = W_a^R W_a^L \underline{b}$. Rewriting $W_a^R W_a^L = W_{aa}^{RL}$ we obtain the Eq. (72). The coefficients of the left and right products of two elements of an eight-dimensional Clifford algebra can be computed using similar matrix representations (for the detailed explanation, please refer to [1]). Leveraging the relationship described in Eq. (65), we can rewrite it in matrix form as:

$$-uxu^{-1} \mapsto -W_{uu^{-1}}^{RL} \, \underline{x} \quad , \tag{73}$$

Therefore, given the Clifford product $\mu = u_1 \cdots u_m$ of m vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and due to the associativity of both Clifford and matrix multiplication, any orthogonal transformation of a multivector x takes the form:

$$(-1)^m \, u_1 \ldots u_m \, x \, (u_1 \ldots u_m)^{-1} \mapsto (-1)^m \, W_{\mu\mu^{-1}}^{RL} \, \underline{x} \quad . \tag{74}$$

2 Clifford neural layers

In this section, we introduce Clifford neural layers, starting from their traditional analogues. We present the Clifford linear layer, the Clifford 2D-convolutional layer and the Clifford 3D-convolutional layer.

2.1 Clifford linear layer

Linear layer. A classic linear layer is an application $l : \mathbb{R}^d \to \mathbb{R}^k$, followed by a component-wise non-linear activation function $\sigma : \mathbb{R} \to \mathbb{R}$, s.t:

$$z_{c'}^{(l)} = \sigma\left(\sum_{c=1}^{d} w_{c,c'}^{(l)} \cdot_{\mathbb{R}} x_c\right) \quad , \qquad \forall c' = 1, ..., k \quad . \tag{75}$$

In other words, the projection of the output onto the j-th direction in the $\mathbb{R}^k$ space is determined by the inner product, computed in $\mathbb{R}^d$ between the j-th learnable weight vector $\boldsymbol{w}_j \in \mathbb{R}^d$ and the input $\boldsymbol{x} \in \mathbb{R}^d$. The activation is subsequently subjected to a component-wise non-linear transformation through the function σ.

Clifford linear layer. Unlike the traditional linear layer, a Clifford linear layer is a mapping $l : Cl_{p,q} \to Cl_{p,q}$ that operates by multiplying the input $x \in Cl_{p,q}$ by a learnable element $w^{(l)} \in Cl_{p,q}$ using the Clifford product.

$$\underline{z}^{(l)} = W^R_{w^{(l)}} \, \underline{x} \ . \tag{76}$$

Any Clifford linear layer learns a multivector $w^{(l)} \in Cl_{p,q}$ and maps the input $x \in Cl_{p,q}$ to the output $z \in Cl_{p,q}$ exploiting the matrix form of the right Clifford product[6], as proposed in Eq. (71).

Inherently capable of potentially representing orthogonal transformations through the Clifford product, a Clifford layer does not explicitly require channel dimensionality expansion or non-linear activation functions to represent any type of geometric transformation of the input. Therefore, unlike traditional methods and the Clifford layers proposed in [1, 7], no non-linear activation functions were included between consecutive layers. Additionally, by default, the d input and k output channels are fixed to 1.

Alternatively, instead of setting $d = k = 1$, channels can be utilized to manipulate spatial dimensionality. This involves arranging the spatial organization of the input data into a single dimension, setting $d = WH$ and $k = W'H'$. This approach simplifies upsampling or downsampling operations without requiring additional parameters or pooling layers:

$$\underline{z}^{(l)}(j) = \sum_{c=1}^{WH} W^R_{w^{(l)}}(c, j) \, \underline{x}(c), \qquad \forall j = 1, \ldots, W'H' \quad . \tag{77}$$

[6] Although it would have been natural to opt for the implementation of $W^{RL}_{ww^{-1}}$ as the weight matrix associated with the layer, in line with the result presented in Eq. (73) we have chosen to follow the approach proposed in [1], which involves associating W^R_w with each Clifford linear layer.

2.2 Clifford 2D convolutional layer

2D-convolutional layer. A convolutional layer is obtained from a linear layer by ensuring sparse connectivity, parameter sharing, locality, and translation equivariance. Multi-channel 2D-convolutional layers are typically used when the input exhibits a two-dimensional spatial structure (i.e., it is an image of size $W \times H$ with d channels). A classical 2D-convolutional layer is a mapping $l : \mathbb{R}^{W\times H\times d} \to \mathbb{R}^{W\times H\times k}$ that convolves the input $\boldsymbol{x} \in \mathbb{R}^{W\times H\times d}$ with a learnable filter bank $\{\boldsymbol{w}_{c'}\}_{c'=1}^{k}$, where each $\boldsymbol{w}_{c'} \in \mathbb{R}^{K\times K\times d}$ is a 2D kernel with an area of $K \times K$ (typically $K = 3$) and d channels:

$$z_{c'}^{(l)}(i,j) = \sigma\left(\sum_{c=1}^{d}\sum_{v,u=-\lfloor\frac{K}{2}\rfloor}^{\lfloor\frac{K}{2}\rfloor} w_{c,c'}^{(l)}(v,u) \cdot_{\mathbb{R}} x_c(i-v,\, j-u)\right), \quad \forall c' = 1,\ldots,k\,. \tag{78}$$

Clifford 2D-convolutional layer. A Clifford 2D-convolutional layer is a mapping $l : Cl_{p,q}^{W\times H} \to Cl_{p,q}^{W\times H}$ that convolves (using the Clifford product $\cdot_{Cl_{p,q}}$, instead of $\cdot_{\mathbb{R}}$) the input $\boldsymbol{x} \in Cl_{(q,p)}^{W\times H}$ with the learnable element $\boldsymbol{w} \in Cl_{(q,p)}^{W\times H}$:

$$\underline{z}^{(l)}(i,j) = \sum_{v,u=-\lfloor\frac{K}{2}\rfloor}^{\lfloor\frac{K}{2}\rfloor} W_{w^{(l)}}^{R}(v,u)\, \underline{x}(i-v,\, j-u)\ . \tag{79}$$

2.3 Clifford 3D convolutional layer

3D-convolutional layer. Multi-channel 3D-convolutional layers are commonly used for inputs with spatial-temporal structure (e.g., video data). They involve mapping an input tensor to an output tensor using a learnable filter bank $\{\boldsymbol{w}_{c'}\}_{c'=1}^{k}$, where each $\boldsymbol{w}_{c'} \in \mathbb{R}^{T\times K\times K\times d}$ is a 3D kernel of volume $T \times K \times K$ across d channels. For each channel $c' = 1,\ldots,k$, the 3D convolutional layer is an application $l : \mathbb{R}^{F\times W\times H\times d} \to \mathbb{R}^{F\times W\times H\times k}$ that processes the input as follows, $\forall c' = 1,\ldots,k$:

$$z_{c'}^{(l)}(t,i,j) = \sigma\left(\sum_{c=1}^{d}\sum_{\tau=-\lfloor\frac{T}{2}\rfloor}^{\lfloor\frac{T}{2}\rfloor}\sum_{v,u=-\lfloor\frac{K}{2}\rfloor}^{\lfloor\frac{K}{2}\rfloor} w_{c,c'}^{(l)}(\tau,v,u) x_c(t-\tau,i-v,j-u)\right)\ . \tag{80}$$

Clifford 3D-convolutional layer. A Clifford 3D-convolutional layer is a mapping $l : Cl_{p,q}^{F\times W\times H} \to Cl_{p,q}^{F\times W\times H}$ that uses the Clifford product to convolve the input $\boldsymbol{x} \in Cl_{p,q}^{W\times H}$ with a single learnable element $\boldsymbol{w}^{(l)} \in Cl_{p,q}^{F\times W\times H}$:

$$\underline{z}^{(l)}(t,i,j) = \sum_{\tau=-\lfloor\frac{T}{2}\rfloor}^{\lfloor\frac{T}{2}\rfloor}\sum_{v,u=-\lfloor\frac{K}{2}\rfloor}^{+\lfloor\frac{K}{2}\rfloor} W_{w^{(l)}}^{R}(\tau,v,u)\, \underline{x}(t-\tau,i-v,j-u)\ . \tag{81}$$

Similarly to Clifford linear layers and Clifford 2D convolutional layers, we employ Clifford 3D convolutional layers without expanding the channel dimensionality or applying non-linear activation functions.

3 Experiments

3.1 Pre-processing, Training and Testing

In this section, we provide a detailed explanation of the pre-processing phase occurring after converting each video into its 3D-motion representation and before the training and testing stages of $\texttt{CliffPhys}_{Cl_{p,q}}$.

Training phase During the training phase, it is necessary to pre-process not only the input video but also its corresponding ground truth. In each training epoch, an input video for the $\texttt{CliffPhys}_{Cl_{p,q}}$ model corresponds to N windows, each containing $S = 399$ frames. These windows are derived from the original video that has been temporally resampled to 20 Hz and spatially resized to 36×36 pixels. A detailed list of the steps involved in video pre-processing are reported in Fig. 1a and summarised below:

- Video spatial resampling: Starting from a spatial resolution of $W \times H$, the motion video is downsampled to a resolution of 36×36 pixels:

$$\boldsymbol{x} \in Cl_{p,q}^{F \times W \times H} \quad \mapsto \quad \boldsymbol{x} \in Cl_{p,q}^{F \times 36 \times 36} \quad . \tag{82}$$

- Video temporal resampling: Beginning with a temporal resolution of F frames at a frame rate of fps, the motion video is downsampled to F' frames at a frame rate of 20Hz:

$$\boldsymbol{x} \in Cl_{p,q}^{F \times 36 \times 36} \quad \mapsto \quad \boldsymbol{x} \in Cl_{p,q}^{F' \times 36 \times 36} \quad . \tag{83}$$

- Temporal Windowing: The resampled version of the data is divided into N non-overlapping windows of $S = 399$ samples (approximately 20 seconds at a 20Hz sampling rate).

$$\boldsymbol{x} \in Cl_{p,q}^{F' \times 36 \times 36} \quad \mapsto \quad [\boldsymbol{x}_1 \in Cl_{p,q}^{399 \times 36 \times 36}, \ ..., \ \boldsymbol{x}_N \in Cl_{p,q}^{399 \times 36 \times 36}] \quad . \tag{84}$$

- Standardization: Considering the n-th window motion signal $(\boldsymbol{\nu}^{(i)}, \boldsymbol{\nu}^{(j)}, \boldsymbol{z})_{n,i,j} \in Cl_{p,q}^{399}$, with spatial coordinates $(i, j) : i, j = 1, \ldots, 36$, the standardization is performed considering the means $\mu^{(z)}, \mu^{(\nu^{(i)})}, \mu^{(\nu^{(j)})}$ and standard deviations $\sigma^{(z)}, \sigma^{(\nu^{(i)})}, \sigma^{(\nu^{(j)})}$ calculated over the COHFACE training set:

$$\widetilde{\boldsymbol{z}}_n(i,j) = \frac{\boldsymbol{z}_n(i,j) - \mu^{(z)}}{\sigma^{(z)}} \qquad \forall i, j = 1, \ldots, 36, \ n = 1, \ldots N \quad , \tag{85}$$

$$\widetilde{\boldsymbol{\nu}}_n^{(i)}(i,j) = \frac{\boldsymbol{\nu}_n^{(i)}(i,j) - \mu^{(\nu^{(i)})}}{\sigma^{(\nu^{(i)})}} \qquad \forall i, j = 1, \ldots, 36, \ n = 1, \ldots N \quad , \tag{86}$$

$$\widetilde{\boldsymbol{\nu}}_n^{(j)}(i,j) = \frac{\boldsymbol{\nu}_n^{(j)}(i,j) - \mu^{(\nu^{(j)})}}{\sigma^{(\nu^{(j)})}} \qquad \forall i,j = 1,\dots,36,\ n = 1,\dots N \quad . \tag{87}$$

The ground truth respiratory signal is pre-processed as follows (Fig. 1b):

- Filtering: the ground truth $\boldsymbol{y} \in \mathbb{R}^F$ is filtered through a second-order Butterworth filter f_b with cutoff frequencies of $[0.1, 0.5]$ Hz.
- GT temporal resampling: starting from a temporal resolution of F samples at a sampling rate of Fs, the ground truth is downsampled to F' samples at a new sampling frequency of 20 Hz:

$$\boldsymbol{y} \in \mathbb{R}^F \quad \mapsto \quad \boldsymbol{y} \in \mathbb{R}^{F'} \quad . \tag{88}$$

- Temporal Windowing: The resampled version of the ground truth is divided into N non-overlapping windows of $S = 399$ samples each (approximately 20 seconds at a 20 Hz sampling rate).

$$\boldsymbol{y} \in \mathbb{R}^{F'} \quad \mapsto \quad [\boldsymbol{y}_1 \in \mathbb{R}^{399},\ \dots,\ \boldsymbol{y}_N \in \mathbb{R}^{399}] \quad . \tag{89}$$

- Standardization: the n-th GT window $\boldsymbol{y}_n \in \mathbb{R}^{399}$ is standardized by considering the mean $\mu^{(y)}$ and standard deviation $\sigma^{(y)}$, calculated over the COHFACE training set:

$$\widetilde{\boldsymbol{y}}_n = \frac{\boldsymbol{y}_n - \mu^{(y)}}{\sigma^{(y)}} \quad . \tag{90}$$

Testing phase For the testing phase, as depicted in Fig. 1d, the video preprocessing is analogous to that conducted during the training phase (Fig. 1a). The N predictions, generated by the model in response to the N windows of $S = 399$ frames provided as input, are concatenated to recreate the original temporal sequence of F instances per video. To compare the respiration per minute (RPM) $\hat{h}$ and h, extracted respectively from the prediction and the ground truth, it is necessary to filter both signals using the f_b filter. The $\hat{h}, h \in \mathbb{R}$ are obtained as the frequencies corresponding to the most prominent peaks in the prediction and GT Power Spectral Densities (PSDs), estimated using the Welch periodogram (cfr. Fig. 3d):

$$\hat{h} = 60 \cdot \underset{f \in [0,10]\text{Hz}}{\arg\max}\ \hat{Y}(f) \quad , \tag{91}$$

$$h = 60 \cdot \underset{f \in [0,10]\text{Hz}}{\arg\max}\ Y(f) \quad , \tag{92}$$

where 10 is the Nyquist frequency. The size of each window for computing the Welch periodogram is $\frac{2F}{3}$ samples. The overlap is 1 second. The number of points at which each one-sided spectrum window is evaluated is fixed at 12000, achieved through the use of zero-padding.

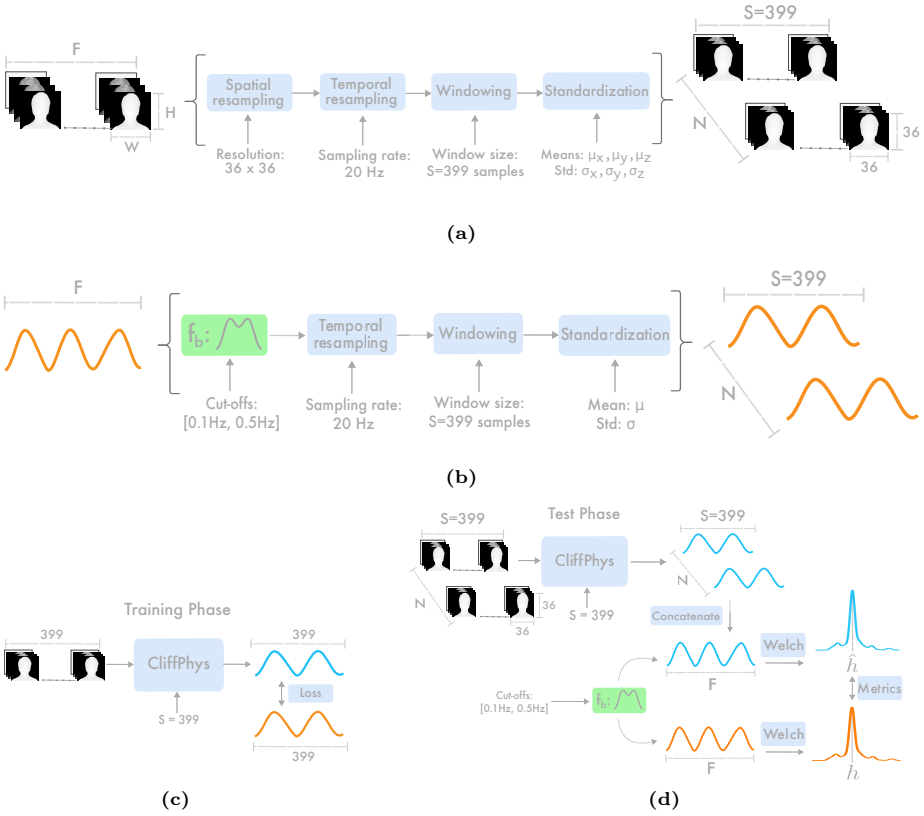

Fig. 1: Pre-processing pipelines implemented during the `CliffPhys`Clp, q' training phase **(c)** and the test phase **(d)**. The input video pre-processing is illustrated in **(a)**, while the corresponding ground truth undergoes pre-processing operations as depicted in **(b)**.

3.2 BP4D+ preprocessing

Some respiratory ground truths in the BP4D+ dataset are significantly affected by artifacts and disturbances. Consequently, only a subset of subjects where considered in the evaluation. The selection procedure closely follows that proposed in [3]. Specifically, the respiratory references were filtered using a second-order Butterworth bandpass filter with cutoff frequencies of 0.1Hz and 0.5 Hz, and then normalized within the amplitude range of $[-1, 1]$. Subsequently, peak and trough detection were conducted to calculate the standard deviation σ_1 of the peak-to-peak time intervals and the standard deviation σ_2 of the heights of the troughs (minima). A signal was labeled as "corrupted" if $\sigma_1 > 1$ or $\sigma_2 > 0.2$, leading to the rejection of the related data. Additionally, any signal with a duration shorter than 20 seconds was discarded. This resulted in 212 accepted signals. In Tab. 1 we report the complete list of considered videos.

Table 1: Table listing the filtered version of the BP4D+ dataset. For every preserved subject, its preserved trials are reported. In the original dataset, trials were named from T1 to T9. In the filtered version the following trials instances were counted: T3: 94, T9: 36, T8: 33, T7: 16, T5: 12, T2: 9, T6: 7, T1: 5. The total number of videos is 212.

Name	Trials
F001	T1, T3, T7
F002	T2, T3, T9
F003	T3, T7
F004	T1, T2, T5
F005	T3, T9
F006	T3, T9
F007	T3, T8
F008	T2, T3
F009	T3, T9
F010	T2, T3, T5, T8
F012	T6
F013	T3
F014	T3
F015	T3
F016	T3, T8, T9
F017	T8
F018	T3, T6, T7
F020	T3, T9
F022	T3, T6, T8
F023	T3, T5, T8
F025	T3, T5
F026	T5
F028	T3
F031	T3
F032	T3, T8
F033	T3
F034	T3, T5, T6, T8
F036	T8
F037	T3, T7
F039	T3
F040	T3, T9
F041	T3, T9
F044	T3
F045	T3, T7
F047	T3
F048	T3, T8
F049	T3, T5, T6
F051	T3, T9
F052	T8, T9

Name	Trials
F054	T3
F055	T3, T8
F056	T3, T7
F057	T3, T8, T9
F058	T3, T9
F059	T2, T3, T9
F060	T8
F061	T3
F062	T3
F063	T3, T9
F064	T3, T7
F065	T3
F066	T3, T7, T9
F067	T3
F068	T1, T3, T6, T8, T9
F069	T3, T9
F070	T3, T9
F071	T3
F072	T7, T9
F073	T3, T9
F074	T3, T9
F076	T7
F077	T7, T8
F078	T3
F079	T3
F080	T9
M001	T3
M002	T3
M003	T3, T8
M004	T8
M005	T1, T3, T8, T9
M006	T2
M007	T3, T7, T8, T9
M008	T2, T3
M009	T3, T8, T9
M010	T7, T9
M011	T3
M012	T3, T9
M013	T3

Name	Trials
M014	T8
M015	T3, T8
M016	T3
M017	T3
M018	T1
M019	T3, T6, T8, T9
M020	T3, T8
M021	T3
M022	T2, T3
M023	T3, T7, T9
M024	T5
M025	T3, T5, T8
M026	T3, T8
M027	T3
M031	T3, T8
M032	T5
M033	T3, T5, T8, T9
M034	T3, T5
M035	T3, T8
M036	T3
M037	T3, T8, T9
M038	T3, T9
M040	T9
M042	T3, T9
M043	T3
M044	T3
M045	T3
M046	T3
M047	T3, T7
M048	T3
M049	T3
M050	T9
M052	T3
M053	T3
M054	T9
M056	T3, T7, T8
M058	T2, T3, T8

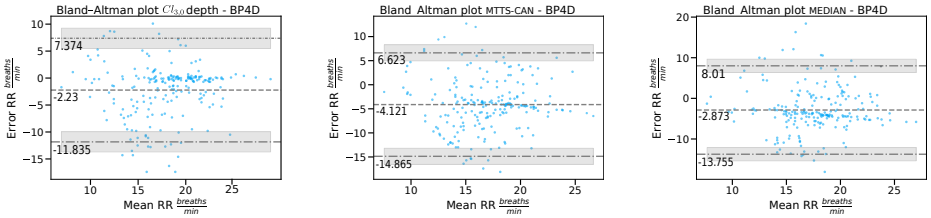

Fig. 2: Bland-Altman plots illustrating the performance of depth-informed $\texttt{CliffPhys}_{Cl_{3,0}}$, in **(a)**, `MTTS-CAN`, in **(b)**, and `Median`, in **(c)**, on the filtered version of the BP4D+ dataset (212 RPM estimates).

3.3 BP4D+ Cross-dataset Results

Bland-Altman analysis allows to quantify the difference between measurements using a graphical method. The Bland-Altman plot represents each pair of values (RPM estimate, RPM reference) as a point connecting the average of the two measurements against their difference, revealing systematic biases: negative errors signify underestimation, while positive errors indicate overestimation. In Figures 2b and 2c, both `MTTS-CAN` and `Median` tend to underestimate the target RPM with a slightly linear trend. Conversely, in Figure 2a for the $\texttt{CliffPhys}_{Cl_{3,0}}$ model, most observations align along the horizontal line corresponding to zero error.

In addition to the qualitative evaluation presented in the main text, we provide the Power Spectral Densities (PSDs) and temporal realizations for two additional subjects from the BP4D+ dataset, namely F068 (trial T9) and M025 (trial T8). Alongside the predictions generated by $\texttt{CliffPhys}_{Cl_{3,0}}$, we also include the performances on individual videos produced using `MTTS-CAN` and the baseline `Median`.

For the first subject (F068, trial T9), it can be observed (Fig. 3a) that the predictions generated by all three approaches exhibit significant frequency content across the entire respiratory range (0.1Hz - 0.5Hz). Although both `MTTS-CAN` and `Median` manage to identify the respiratory motion, only $\texttt{CliffPhys}_{Cl_{3,0}}$ recognizes it as predominant. Conversely, the two other methods identify a slightly higher frequency artifact as the main motion. The corresponding temporal signals are depicted in Fig. 3b.

For the second subject (M025, trial T8), Fig. 3c illustrates how the PSD of the estimation produced by the $\texttt{CliffPhys}_{Cl_{3,0}}$ model closely overlaps with the ground truth (GT) PSD. The remaining methods fail to identify the respiratory motion as clearly: the `MTTS-CAN` prediction presents four peaks corresponding to potential respiratory frequencies, while `Median` interprets a frequency present in the GT as an artifact.

The images provided demonstrate how $\texttt{CliffPhys}_{Cl_{3,0}}$ is capable of making accurate predictions regardless of the specific trial (trial T9 for F068, whereas T8 was the trial for the qualitative example presented in the main text) and

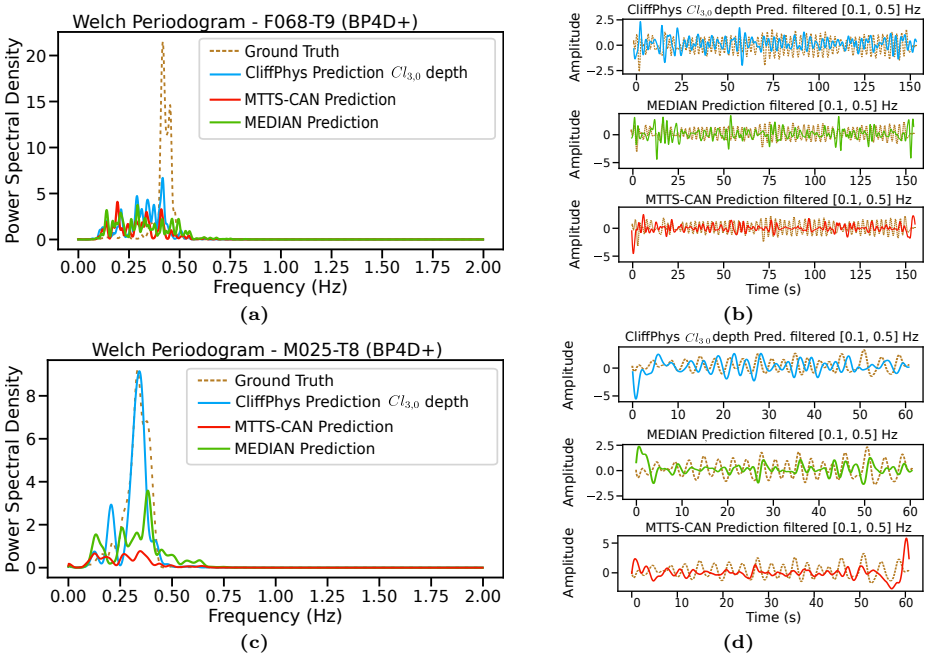

Fig. 3: Examples of PSDs, **(a)**, **(c)**, and respiratory prediction signals, **(b)**, **(d)**, for subjects F068-T9 and M025-T8 in the BP4D+ dataset, displayed for MTTS-CAN (red), MEDIAN (green), and depth-informed $\texttt{CliffPhys}_{Cl_{3,0}}$ (light-blue), compared to the GTs (gold).

regardless of the participant's gender (here, we present the prediction for the male subject M025, whereas the main text showcased the qualitative result for a female subject). The respective temporal signals are displayed in Fig. 3d.

A comprehensive statistical analysis has been conducted considering the predictions generated, in the cross-dataset evaluation task, by all models. In particular, let us consider the absolute distances $|\hat{h}_m^{(k)} - h_m|$ between the RPMs produced by the k-th model, with $k = 1, \ldots, 17$, for the subjects $m = 1, \ldots, 212$ from filtered BP4D+. In this context, the 212 RPMs $\hat{\boldsymbol{h}}^{(k)}$, predicted through the 17 different methods, are understood as realizations of different populations, ideally described by as many probability distributions. In other words, 17 different ways of modeling the respiratory phenomenon, more or less faithful to the original distribution, are evaluated considering $\hat{h}_m^{(k)}$ as realizations and $|\hat{h}_m^{(k)} - h_m|$ as a way to assess the discrepancy between prediction and true label.

Given that there are more than two populations and some of them are not normal (Shapiro-Wilk test, $p < 0.05$), the non-parametric Friedman test is employed as an omnibus test to ascertain significant differences among median values of the populations. The post-hoc Nemenyi test is then applied to identify significant differences. Tab. 2 reports results across all populations and samples.

Model	MED	MAD	CI	γ	Magnitude
`CliffPhys`$_{Cl_{3,0}}$ (w/ depth)	1.700	1.600	[0.600, 3.800]	0.000	negligible
`CliffPhys`$_{Cl_{0,3}}$ (w/ depth)	1.850	1.750	[0.600, 4.400]	-0.060	negligible
`CliffPhys`$_{Cl_{0,2}}$ (w/o depth)	2.250	2.050	[0.600, 4.800]	-0.202	small
`CliffPhys`$_{Cl_{0,2}}$ (w/ depth)	2.300	2.100	[0.600, 4.300]	-0.217	small
`CliffPhys`$_{Cl_{2,0}}$ (w/o depth)	1.900	1.750	[0.700, 4.200]	-0.080	negligible
`CliffPhys`$_{Cl_{2,0}}$ (w/ depth)	1.600	1.500	[0.500, 5.100]	0.043	negligible
`CliffPhys`$_{Cl_{3,0}}$ (w/o depth)	2.200	2.000	[0.800, 4.800]	-0.186	negligible
`CliffPhys`$_{Cl_{0,3}}$ (w/o depth)	2.700	2.500	[1.100, 5.000]	-0.321	small
`CliffPhys`$_{Cl_{3,0}}$ (R,G,B)	4.200	3.400	[2.100, 7.000]	-0.635	medium
`Median`	4.200	1.600	[3.500, 5.200]	-1.054	large
3D conv (R,G,B)	5.300	2.400	[3.700, 6.300]	-1.191	large
3D conv (w/ depth)	4.900	3.300	[3.000, 7.000]	-0.832	large
`PhysFormer`	4.300	2.300	[3.100, 6.100]	-0.885	large
`Big-Small`	4.750	1.950	[3.800, 5.800]	-1.153	large
`MTTS-CAN`	4.600	1.900	[3.900, 5.900]	-1.114	large
`Profile1D`	4.500	2.100	[3.800, 6.100]	-1.012	large
`ContrastPhys`	8.500	3.300	[6.700, 10.300]	-1.769	large

Table 2: Summary of populations

For each model we report the median MAE (MED), median absolute deviation (MAD), confidence interval (CI), effect size (Akinshin's gamma, γ) and its magnitude. Differences between populations are considered significant if the difference in mean rank exceeds the critical distance CD=1.696 of the Nemenyi test.

The null hypothesis of the Friedman test, indicating no difference in central tendency among populations, is rejected ($p < 0.05$). Thus, it is concluded that there exists a statistically significant difference in median values among the populations. According to the post-hoc Nemenyi test, all models from the `CliffPhys`$_{Cl_{p,q}}$ family deliver significantly better results if compared with other

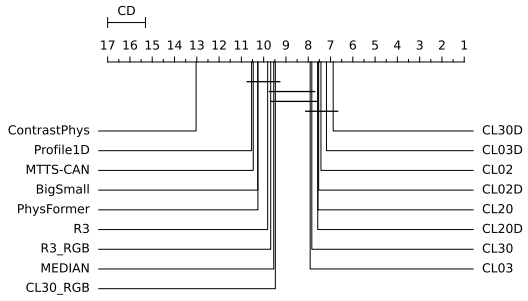

Fig. 4: Critical difference (CD) diagram [4] ranking all the estimation methods (using MAE as the evaluation metric).

models and baselines, with models ingesting depth information reaching higher ranks.

References

1. Brandstetter, J., Berg, R.v.d., Welling, M., Gupta, J.K.: Clifford neural layers for pde modeling. arXiv preprint arXiv:2209.04934 (2022) 8, 9
2. Dorst, L., Fontijne, D., Mann, S.: Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2009) 1
3. Fiedler, M.A., Rapczyński, M., Al-Hamadi, A.: Fusion-based approach for respiratory rate recognition from facial video images. IEEE Access **8**, 130036–130047 (2020) 13
4. Herbold, S.: Autorank: A python package for automated ranking of classifiers. Journal of Open Source Software **5**(48), 2173 (2020). https://doi.org/10.21105/joss.02173, https://doi.org/10.21105/joss.02173 17
5. Lounesto, P.: Clifford algebras and spinors. In: Clifford Algebras and Their Applications in Mathematical Physics, pp. 25–37. Springer (2001) 1, 4
6. Macdonald, A.: Linear and geometric algebra. Alan Macdonald Nottingham (2010) 1, 4
7. Ruhe, D., Gupta, J.K., De Keninck, S., Welling, M., Brandstetter, J.: Geometric clifford algebra networks. arXiv preprint arXiv:2302.06594 (2023) 9